

Estimation of mean form and mean form difference under elliptical laws

Francisco J. Caro-Lopera *

Departament of Basic Sciences
Universidad de Medellín
Medellín, Colombia
E-mail: fjcaro@udem.edu.co

José A. Díaz-García

Universidad Autónoma Agraria Antonio Narro
Calzada Antonio Narro 1923, Col. Buenavista
25315 Saltillo, Coahuila, México
E-mail: jadiaz@uaaan.mx

Abstract

Some ideas studied by Lele (1993), under a Gaussian perturbation model, are generalised in the setting of matrix multivariate elliptical distributions. In particular, several inaccuracies in the published statistical perturbation model are revised. In addition, a number of aspects about identifiability and estimability are also considered. Instead of using the Euclidean distance matrix for proposing consistent estimates, this paper determines exact formulae for the moments of matrix $\mathbf{B} = \mathbf{X}^c (\mathbf{X}^c)^T$, where \mathbf{X}^c is the centered landmarks matrix. Consistent estimation of mean form difference under elliptical laws is also studied. Finally, the main results of the paper and some methodologies for selecting models and hypothesis testing are applied to a real landmark data. comparing correlation shape structure is proposed and applied in handwritten differentiation.

1 Introduction

Statistical theory of shape has emerged as one of the most versatile techniques of classification and comparison of “objects” in a number of disciplines. By its theoretical nature, the matrix multivariate distribution analysis fits very well into the shape analysis, but at the same time have involved strong open problems on estimation of location and scale population parameters based on the exact distributions, forcing the application of several less robust approaches, which were considered appropriate at first, but later received important critics from different experts on morphometrics and related fields, see Lele (1993).

Among the addressed lacks we can cite the use of asymptotic distributions, tangent plane inference, isotropic models, Gaussian assumptions, and procrustes theory; for details of such techniques see Dryden and Mardia (1998) and the references therein.

*Corresponding author

Key words. Coordinate free approach, non-central singular Pseudo-Wishart distribution, statistical shape theory, matrix multivariate elliptical distribution, matrix multivariate gaussian distribution.
2000 Mathematical Subject Classification. 62E15; 62E05; 62H12; 62H30; 62H35

Now, some attempts have been published recently avoiding the above restrictions and considering inference via likelihood function using the exact shape distributions, the new theory was termed generalised shape theory by finding the exact shape densities indexed by families of distributions of elliptical contours. According to the geometrical filters on shape, the resulting exact invariant distributions are expressed in terms of series of functions termed Jack polynomials, which were uncomputable for decades, and only recently with the works of Koev and Edelman (2006), the individual polynomials could be computed but series of them have involved serious problems in inference of population parameters via likelihood method. A number of approaches with a meaningful computational success in the context of the classical Gaussian and elliptical models are given as follows: via QR decomposition, see Goodall and Mardia (1993) and Díaz-García and Caro-Lopera (2014), singular value decompositions, see Le and Kendall (1993), Goodall (1991), Díaz-García *et al.* (2003), Díaz-García and Caro-Lopera (2012a) and Díaz-García and Caro-Lopera (2012b)), affine see Goodall and Mardia (1993), Díaz-García *et al.* (2003), Caro-Lopera *et al.* (2009), Caro-Lopera and Díaz-García (2012) and Caro-Lopera *et al.* (2014)), and Pseudo-Wishart, see Díaz-García and Caro-Lopera (2013). However, a feasible approach dealing with computable exact densities and a likelihood function based on polynomials of very low degree was published recently, letting robust estimation on location and scale population parameters very accurate; it models shapes under certain conditions via affine transformation, which means that it removes from objects, any geometrical information of rotation, translation, scaling and uniform shear. Meanwhile, the similarity (Euclidean) transformations via QR, SVD, Pseudo-Wishart (invariant under rotation, translation, scaling) capture the attention of most of the users of shape theory and is the source of the main critics.

Under Euclidean transformations the shape distributions are extremely difficult to compute and then the associated inference, it forces the use of isotropic models, an assumption which is unrealistic in biology, for example, since it says that landmarks vary independently of each other along different axes but are correlated along a fixed axis. In fact, this isotropy assumption is very common in literature, leaving the problem of testing solely whether shapes are equal; but for biologists, for example, they want to identify the correlation structure of landmarks and the structures of shape which absorbs the meaningful differences. Moreover, estimation of a full covariance structure would be the desirable result, because, correlation among landmarks and axis is important, but the correlation among objects in the sample should provide a complete comprehension of the involved populations, see Lele and Richtsmeier (1990)

Instead of estimation via likelihood method, some authors have proposed, the Gaussian case, the method-of-moments estimators of the mean form and the variance-covariance structure which are consistent and simple to compute, see for example Lele and Richtsmeier (1991), Lele (1993), Richtsmeier *et al.* (2002) and the references therein. In fact, the technique was set as a critics of generalised procrustes analysis, by proving that application of the last analysis yield inconsistent estimators of the mean form, mean shape, and variance-covariance structure, and then all the statistical inference procedures can produce inaccurate results. Walker (2001) recently reiterated the conclusions of Lele (1993) by reporting the inability of Procrustes methods to estimate the correct variance-covariance structure and the associated implications for statistical inference. This aspect, is crucial because Procrustes analysis is one of the most common method of estimation in several fields such as morphometrics.

Given the computation problems of maximum likelihood estimators, the method-of-moments estimators emerges as one of the promissory techniques in shape theory, however some open problems must be studied deeply. For example, some inaccuracies of this model presented in Lele (1993), assuming a matrix multivariate Gaussian distribution must be nuanced first, and second, the method should allow non Gaussian samples, a realistic and

very common problem in morphometrics and the usual applied areas for shape, a suitable solution comes from families of elliptical contoured distributions, which exhibit lighter or heavier tails, or greater or less kurtosis than the Gaussian model. Setting generalised shape theory also must include criteria for selecting models and hypothesis testing, in order to provide an integrating theory suitable to be applied in meaningful scenarios.

Clarifying the inaccuracies of Lele (1993) ideas, and their connection with some theoretical studies by Magnus and Neudecker (1979), Muirhead (1982) and Díaz-García (1994) should give a unified theory setting the isolated Gaussian approach into the general framework of the existing generalised matrix multivariate elliptical shape theory.

Thus, estimation of mean form and mean form difference under elliptical laws is placed in this work as follows: Section 2 clarifies some results of the published Gaussian case and propose the generalisation in the context of matrix multivariate elliptical distributions, it includes some properties of matrix multivariate elliptical distribution, identifiability and estimability of the parameters of interest, the perturbation model under a matrix multivariate elliptical distribution, and invariance and nuisance parameters. Then Section 3 studies the consistent estimation of the population parameters under dependence and independence and provides exact formulae for the moments estimators. Section 4 provides a consistent estimation for a general non-negative definite correlation matrix. The analysis also includes extensions to elliptical models of form difference under the perspective of Euclidean Distance Matrix, see Section 5. Finally, a complete example collecting the main results of the paper and proposing some selecting model criteria, is proposed in Section 6.

2 Preliminary results

In this section we review some notation and distributional results. Also, the statistical model to be used throughout the paper, is established and analysed. In particular some inaccuracies of this model presented in Lele (1993), assuming a matrix multivariate Gaussian distribution are corrected and then is generalised to the case where an matrix multivariate elliptical distribution is assumed.

2.1 Matrix multivariate elliptical distribution

A detailed discussion of the matrix multivariate elliptical distribution can be found for example in Fang and Zhang (1990) and Gupta and Varga (1993), among many others.

Remark 2.1. For matrix multivariate Gaussian and elliptical distributions, traditionally are used two forms for establish that a random matrix \mathbf{Y} has one of these distribution. For example in matrix multivariate Gaussian case, this fact is written as

$$\mathbf{Y} \sim \mathcal{N}_{K \times D}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Theta}),$$

see Arnold (1981), Dutilleul (1999) and Fang and Zhang (1990) among many others authors. However, as is study in Lele (1993), Dutilleul (1999) among others, in general the parameters $\boldsymbol{\Sigma}$ and $\boldsymbol{\Theta}$ are not identifiable one-by-one, but $\boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}$ or $\boldsymbol{\Theta} \otimes \boldsymbol{\Sigma}$ are identifiable. Here \otimes denotes the usual Kronecker product. In addition, given that $\text{Cov}(\text{vec } \mathbf{Y}) = \boldsymbol{\Theta} \otimes \boldsymbol{\Sigma}$, and $\text{Cov}(\text{vec } \mathbf{Y}^T) = \boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}$, many other authors use the notation

$$\mathbf{Y} \sim \mathcal{N}_{K \times D}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}),$$

where “vec” denotes the vectorisation operator, see Muirhead (1982) and Gupta and Varga (1993). Analogous situation is present for matrix multivariate elliptical distributions. We shall use this last notation.

Definition 2.1. It is say that \mathbf{Y} has a matrix multivariate elliptical distribution, with location parameter matrix $\boldsymbol{\mu} \in \Re^{K \times D}$ and scala parameter matrix $\boldsymbol{\Sigma} \otimes \boldsymbol{\Theta} \in \Re^{KD \times KD}$; where $\boldsymbol{\Sigma}$ is a definite positive matrix, $\boldsymbol{\Sigma} > 0$ and $\boldsymbol{\Theta} > 0$, with $\boldsymbol{\Sigma} \in \Re^{K \times K}$ and $\boldsymbol{\Theta} \in \Re^{D \times D}$, if its density function with respect to Lebesgue measure is given by

$$dF_{\mathbf{Y}}(\mathbf{Y}) = |\boldsymbol{\Sigma}|^{-D/2} |\boldsymbol{\Theta}|^{-K/2} h[\text{tr } \boldsymbol{\Theta}^{-1}(\mathbf{Y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu})](d\mathbf{Y}), \quad (1)$$

where the function $h : \Re \rightarrow [0, \infty)$ is such that $\int_0^\infty u^{KD/2-1} h(u) du < \infty$. The function h is termed the *density generator*. Its characteristic function is given by

$$\psi_{\mathbf{Y}}(\mathbf{T}) = \text{etr}(i\boldsymbol{\mu}^T \mathbf{T}) \phi(\text{tr } \mathbf{T} \boldsymbol{\Theta} \mathbf{T}^T \boldsymbol{\Sigma}), \quad (2)$$

with $i = \sqrt{-1}$, $\phi : [0, \infty) \rightarrow \Re$ and $\text{etr}(\cdot) = \exp(\text{tr}(\cdot))$. This fact is denoted as $\mathbf{Y} \sim \mathcal{E}_{K \times D}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}, h)$. In addition, observe that the characteristic function exist still when $\boldsymbol{\Sigma}$ and/or $\boldsymbol{\Theta}$ are semidefinite positive matrices; in such case it say that \mathbf{Y} has a *singular matrix multivariate elliptical distribution*, see Remark 2.2 below.

In addition, note that $\text{Cov}(\text{vec } \mathbf{Y}) = c_0 \boldsymbol{\Theta} \otimes \boldsymbol{\Sigma}$, and $\text{Cov}(\text{vec } \mathbf{Y}^T) = c_0 \boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}$ where $c_0 = -2\phi'(0)$,

$$\phi'(0) = \left. \frac{d\phi(t^2)}{dt} \right|_{t=0}.$$

see (Fang and Zhang, 1990, Theorm 2.6.5, p. 62) and (Gupta and Varga, 1993, Corollary 3.2.1.1, p. 94 and Theorem 2.4.1, p. 33).

Is easy to see that if

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_{(1)}^T \\ \mathbf{Y}_{(2)}^T \\ \vdots \\ \mathbf{Y}_{(k)}^T \end{pmatrix} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_D), \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{(1)}^T \\ \boldsymbol{\mu}_{(2)}^T \\ \vdots \\ \boldsymbol{\mu}_{(k)}^T \end{pmatrix} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_D),$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1K} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{K1} & \sigma_{K2} & \cdots & \sigma_{KK} \end{pmatrix} \text{ and } \boldsymbol{\Theta} = \begin{pmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1D} \\ \theta_{21} & \theta_{22} & \cdots & \theta_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{D1} & \theta_{D2} & \cdots & \theta_{kD} \end{pmatrix}.$$

Then from (Fang and Zhang, 1990)

1. $\mathbf{Y}_{(i)} \sim \mathcal{E}_D(\boldsymbol{\mu}_{(i)}, \sigma_{ii} \boldsymbol{\Theta}, h)$, $i = 1, 2, \dots, K$,
2. $\mathbf{Y}_j \sim \mathcal{E}_K(\boldsymbol{\mu}_j, \theta_{jj} \boldsymbol{\Sigma}, h)$, $j = 1, 2, \dots, D$,

this is,

1. $\text{Cov}(\mathbf{Y}_{(i)}) = c_0 \sigma_{ii} \boldsymbol{\Theta}$, $i = 1, 2, \dots, K$,
2. $\text{Cov}(\mathbf{Y}_j) = c_0 \theta_{jj} \boldsymbol{\Sigma}$, $j = 1, 2, \dots, D$.

These two last affirmations are incorrect stated in Lele (1993) in the context of the perturbation model. For this asseveration observe that this class of matrix multivariate elliptical distributions includes Gaussian, contaminated Gaussian, Pearson type II and VII, Kotz, Jensen-Logistic, power exponential and Bessel distributions, among others; these distributions have tails that are more or less weighted, and/or present a greater or smaller degree of kurtosis than the Gaussian distribution.

2.2 Identifiability and estimability of the parameters of interest

Now some aspects about the identifiability and estimability of the parameters $(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Theta})$ are studied.

Note that the density (1) can be write as, see (Muirhead , 1982, p. 79) and (Gupta and Varga, 1993, Theorem 2.1.1, p. 20),

$$dF_{\text{vec } \mathbf{Y}^T}(\text{vec } \mathbf{Y}^T) = |\boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}|^{-1/2} h[\text{vec}^T(\mathbf{Y} - \boldsymbol{\mu})(\boldsymbol{\Sigma} \otimes \boldsymbol{\Theta})^{-1} \text{vec}(\mathbf{Y} - \boldsymbol{\mu})](d \text{vec } \mathbf{Y}^T), \quad (3)$$

using the fact that $\text{vec}^T \mathbf{X}(\mathbf{D}\mathbf{B} \otimes \mathbf{C}^T) \text{vec } \mathbf{X} = \text{tr}(\mathbf{B}\mathbf{X}^T \mathbf{C}\mathbf{X}\mathbf{D})$, with $\text{vec}^T \mathbf{X} \equiv (\text{vec } \mathbf{X})^T$, and that for $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{B} \in \mathbb{R}^{m \times m}$, then $|\mathbf{A}|^m |\mathbf{B}|^n = |\mathbf{A} \otimes \mathbf{B}|$, see (Muirhead , 1982, Section 2.2, pp. 72-76) and (Fang and Zhang, 1990, Section 1.4, pp. 11-13). Then, denoting $\text{vec } \mathbf{Y}^T = \mathbf{y} \in \mathbb{R}^{KD}$ and $\boldsymbol{\Sigma} \otimes \boldsymbol{\Theta} = \boldsymbol{\Xi}$, the density (3) define the distribution of the vector \mathbf{y} ; moreover, $\mathbf{y} \sim \mathcal{E}_{KD}(\text{vec } \boldsymbol{\mu}, \boldsymbol{\Xi}, h)$.

Now, assume that our data consist of a sample of matrices of size n from a given population, namely $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n$, and define the random matrix

$$\mathbb{Y} = (\text{vec } \mathbf{Y}_1^T, \text{vec } \mathbf{Y}_2^T, \dots, \text{vec } \mathbf{Y}_n^T)^T \in \mathbb{R}^{n \times KD}.$$

From Díaz-García (1994), assuming that $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n$ be independent, the density function of \mathbb{Y} admit the expression

$$dF_{\mathbb{Y}}(\mathbb{Y}) = |\boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}|^{-n/2} h[\text{tr}(\mathbb{Y} - \mathbb{M})(\boldsymbol{\Sigma} \otimes \boldsymbol{\Theta})^{-1}(\mathbb{Y} - \mathbb{M})^T](d\mathbb{Y}), \quad (4)$$

where

$$\mathbb{M} = \mathbf{1}_n \text{vec}^T \boldsymbol{\mu} \in \mathbb{R}^{n \times KD},$$

and $\mathbf{1}_n = (1, 1, \dots, 1)^T \in \mathbb{R}^n$; this is, $\mathbb{Y} \sim \mathcal{E}_{n \times KD}(\mathbb{M}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Theta} \otimes \mathbf{I}_n, h)$. Thus, taking $p = KD$ in (Fang and Zhang, 1990, Theorem 4.1.1, p.129), and given that $KD < n$ and $h(\cdot)$ being nonincreasing and continuous, we have that the *maximum likelihood estimate* of $(\text{vec } \boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Theta})$ is

$$\left(\widetilde{\text{vec } \boldsymbol{\mu}}, \widetilde{\boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}} \right) = (\bar{\mathbf{y}}, \lambda_{\max} \mathbf{S}),$$

where λ_{\max} is the critical point where the function $h^*(\lambda)$ has its maximum, with

$$h^*(\lambda) = \lambda^{-KDn/2} h(KD/\lambda),$$

$$\bar{\mathbf{y}} = \frac{1}{n} \mathbb{Y}^T \mathbf{1}_n \in \mathbb{R}^{KD}, \text{ and } \mathbf{S} = \mathbb{Y}^T \mathbf{H}_n \mathbb{Y} \in \mathbb{R}^{KD \times KD}$$

where $\mathbf{H}_n = \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T$, defines an orthogonal projection, this is, $\mathbf{H}_n = \mathbf{H}_n^T = \mathbf{H}_n^2$. Or alternatively

$$\bar{\mathbf{y}} = \frac{1}{n} \sum_{i=1}^n \text{vec } \mathbf{Y}_i^T, \text{ and } \mathbf{S} = \sum_{i=1}^n (\text{vec } \mathbf{Y}_i^T - \bar{\mathbf{y}})(\text{vec } \mathbf{Y}_i^T - \bar{\mathbf{y}})^T,$$

from where the estimator of $\boldsymbol{\mu}$ is

$$\tilde{\boldsymbol{\mu}} = \bar{\mathbf{Y}} = \frac{1}{n} \sum_{i=1}^n \mathbf{Y}_i.$$

From (Fang and Zhang, 1990, Section 4.3), several properties of the maximum likelihood estimators $\tilde{\boldsymbol{\mu}}$ and $\widetilde{\boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}}$ are obtained as: *sufficiency, completeness, consistency and unbiasedness*. Specifically, for \mathbb{Y} with the finite 2nd moment and $h(\cdot)$ be nonincreasing and continuous,

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{Y}} \quad \text{and} \quad \widehat{\boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}} = \frac{1}{2(1-n)\psi'(0)} \mathbf{S},$$

are *unbiased estimators* of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}$.

Remark 2.2. Observe that when $\mathbf{Y} \sim \mathcal{E}_{K \times D}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}, h)$ and its columns and/or its rows are dependent linearly, is say that \mathbf{Y} has a *singular matrix multivariate elliptical distribution*. Then \mathbf{Y} has density with respect to *Hasusdorff measure*. Moreover, such dependent linearly among its columns or its rows is archived in the rank of $\boldsymbol{\Sigma}$ and/or $\boldsymbol{\Theta}$ matrices and is denoted as: $\mathbf{Y} \sim \mathcal{E}_{K \times D}^{s,r}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}, h)$, where $s = \text{rank}(\boldsymbol{\Sigma}) \leq K$ and $r = \text{rank}(\boldsymbol{\Theta}) \leq D$, see (Gupta and Varga, 1993, Definition 2.1.1, p. 19), Díaz-García and González-Farías (2005) and Díaz-García and Gutiérrez-Jáimez (2006). As in the singular matrix multivariate Gaussian case, the maximum likelihood estimators in singular matrix multivariate elliptical models remain valid, see Khatri (1968) and (Rao, 1973, Section 8a.5, pp.528-532).

2.3 Perturbation model under a matrix multivariate elliptical distribution

Let $\mathbf{X} \in \mathbb{R}^{K \times D}$ a random matrix representing the geometrical figure comprising K landmark, or labeled, points of dimension D , such that $K > D$. This matrix \mathbf{X} is termed *landmark coordinate matrix*, see Lele (1993).

Let $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ be a independently sample of size n of landmark coordinate matrices $\mathbf{X}_i \in \mathbb{R}^{K \times D}$, $i = 1, 2, \dots, n$, from a given population.

The statistical model to be considered in this work is a generalisation of the perturbation model used by Lele (1993) among others authors. Let $\boldsymbol{\mu} \in \mathbb{R}^{K \times D}$ corresponding to the mean form. Let

$$\mathbf{X}_i = (\boldsymbol{\mu} + \mathbf{E}_i)\boldsymbol{\Gamma}_i + \mathbf{t}_i, \quad i = 1, 2, \dots, n, \quad (5)$$

where $\mathbf{E}_i \sim \mathcal{E}_{K \times D}(\mathbf{0}, \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Sigma}_D, h)$, $\boldsymbol{\Gamma}_i \in \mathbb{R}^{D \times D}$ are orthogonal matrices representing rotation and/or reflection of $(\boldsymbol{\mu} + \mathbf{E}_i)$, and $\mathbf{t}_i \in \mathbb{R}^{K \times D}$ are matrices such that $\mathbf{t}_i = \mathbf{1}_k \mathbf{a}_i^T$ representing translation, for some $\mathbf{a}_i \in \mathbb{R}^D$. From (Fang and Zhang, 1990, eq. (3.3.10), p. 103) or (Gupta and Varga, 1993, Theorem 2.1.2, p. 20) we have

$$\mathbf{X}_i \sim \mathcal{E}_{K \times D}(\boldsymbol{\mu}\boldsymbol{\Gamma}_i + \mathbf{t}_i, \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Gamma}_i^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_i, h), \quad i = 1, 2, \dots, n. \quad (6)$$

Parameters of interest are $(\boldsymbol{\mu}, \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Sigma}_D)$ and $(\boldsymbol{\Gamma}_i^T, \mathbf{t}_i)$ $i = 1, 2, \dots, n$ are the nuisance parameters. An detail explained of this perturbation model is given in Lele (1993) among others.

Alternatively, the model (5) can be write as:

$$\text{vec } \mathbb{X}^T = \text{diag}(\mathbb{G}) \text{vec}(\mathbb{M} + \mathbb{E})^T + \text{vec } \mathbb{T}^T,$$

with

$$\text{diag}(\mathbb{G}) = \begin{pmatrix} \mathbf{I}_K \otimes \boldsymbol{\Gamma}_1^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_K \otimes \boldsymbol{\Gamma}_2^T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_K \otimes \boldsymbol{\Gamma}_n^T \end{pmatrix}$$

or the model (5) may be rewritten in the form

$$\mathbb{X} = \text{diag}(\mathbb{M} + \mathbb{E})\mathbb{G}^T + \mathbb{T},$$

with

$$\text{diag}(\mathbb{M} + \mathbb{E}) = \begin{pmatrix} \text{vec}^T(\boldsymbol{\mu} + \mathbf{E}_1)^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \text{vec}^T(\boldsymbol{\mu} + \mathbf{E}_2)^T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \text{vec}^T(\boldsymbol{\mu} + \mathbf{E}_n)^T \end{pmatrix}$$

where

$$\mathbb{X} = \begin{pmatrix} \text{vec}^T \mathbf{X}_1^T \\ \text{vec}^T \mathbf{X}_2^T \\ \vdots \\ \text{vec}^T \mathbf{X}_n^T \end{pmatrix}, \quad \mathbb{M} = \mathbf{1}_n \text{vec}^T \boldsymbol{\mu}^T, \quad \mathbb{E} = \begin{pmatrix} \text{vec}^T \mathbf{E}_1^T \\ \text{vec}^T \mathbf{E}_2^T \\ \vdots \\ \text{vec}^T \mathbf{E}_n^T \end{pmatrix} \quad \mathbb{T} = \begin{pmatrix} \text{vec}^T \mathbf{t}_1^T \\ \text{vec}^T \mathbf{t}_2^T \\ \vdots \\ \text{vec}^T \mathbf{t}_n^T \end{pmatrix},$$

and $\mathbb{G} = (\mathbf{I}_K \otimes \boldsymbol{\Gamma}_1^T | \mathbf{I}_K \otimes \boldsymbol{\Gamma}_2^T | \cdots | \mathbf{I}_K \otimes \boldsymbol{\Gamma}_n^T)$ where

$$\mathbb{E} \sim \mathcal{E}_{n \times KD}(\mathbf{0}, \mathbf{I}_n \otimes \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Sigma}_D, h),$$

or

$$\text{vec} \mathbb{E}^T \sim \mathcal{E}_{nKD}(\text{vec} \mathbf{0}, \mathbf{I}_n \otimes \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Sigma}_D, h),$$

Hence

$$\text{vec} \mathbb{X}^T \sim \mathcal{E}_{nKD}(\text{diag}(\mathbb{G}) \text{vec} \mathbb{M}^T + \text{vec} \mathbb{T}^T, \text{diag}(\mathbb{G})(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Sigma}_D) \text{diag}(\mathbb{G})^T, h).$$

Note that, recalling that for \mathbf{x} and \mathbf{y} vectors, $\text{vec}(\mathbf{y}\mathbf{x}^T) = \mathbf{x} \otimes \mathbf{y}$, then

$$\begin{aligned} \text{diag}(\mathbb{G}) \text{vec} \mathbb{M}^T &= \begin{pmatrix} \mathbf{I}_K \otimes \boldsymbol{\Gamma}_1^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_K \otimes \boldsymbol{\Gamma}_2^T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_K \otimes \boldsymbol{\Gamma}_n^T \end{pmatrix} (\mathbf{1}_n \otimes \text{vec} \boldsymbol{\mu}^T) \\ &= \begin{pmatrix} (\mathbf{I}_K \otimes \boldsymbol{\Gamma}_1^T) \text{vec} \boldsymbol{\mu}^T \\ (\mathbf{I}_K \otimes \boldsymbol{\Gamma}_2^T) \text{vec} \boldsymbol{\mu}^T \\ \vdots \\ (\mathbf{I}_K \otimes \boldsymbol{\Gamma}_n^T) \text{vec} \boldsymbol{\mu}^T \end{pmatrix} \\ &= \begin{pmatrix} \text{vec}(\boldsymbol{\mu}\boldsymbol{\Gamma}_1)^T \\ \text{vec}(\boldsymbol{\mu}\boldsymbol{\Gamma}_2)^T \\ \vdots \\ \text{vec}(\boldsymbol{\mu}\boldsymbol{\Gamma}_n)^T \end{pmatrix}, \end{aligned}$$

and $\text{diag}(\mathbb{G})(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Sigma}_D) \text{diag}(\mathbb{G})^T$ is

$$\begin{aligned} &= \begin{pmatrix} \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Gamma}_1^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Gamma}_2^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Gamma}_n^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_n \end{pmatrix} \\ &= \sum_{i=1}^n \mathbf{E}_{ii}^n \otimes \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Gamma}_i^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_i, \end{aligned}$$

where if \mathbf{e}_i^n the i th column unit vector of order n , then $\mathbf{E}_{ii}^n = \mathbf{e}_i^n (\mathbf{e}_i^n)^T$.

Finally observe that

$$E(\text{vec} \mathbb{X}^T) = \begin{pmatrix} \text{vec}(\boldsymbol{\mu}\boldsymbol{\Gamma}_1)^T \\ \text{vec}(\boldsymbol{\mu}\boldsymbol{\Gamma}_2)^T \\ \vdots \\ \text{vec}(\boldsymbol{\mu}\boldsymbol{\Gamma}_n)^T \end{pmatrix} + \text{vec} \mathbb{T}^T = \sum_{i=1}^n \mathbf{e}_i^n \otimes (\text{vec}(\boldsymbol{\mu}\boldsymbol{\Gamma}_i)^T + \text{vec} \mathbf{t}_i^T)$$

Then

$$E(\mathbb{X}) = \sum_{i=1}^n \mathbf{e}_i^n (\text{vec}(\boldsymbol{\mu}\boldsymbol{\Gamma}_i)^T + \text{vec} \mathbf{t}_i^T)^T.$$

Therefore

$$\mathbb{X} \sim \mathcal{E}_{n \times KD} \left(\sum_{i=1}^n \mathbf{e}_i^n (\text{vec}(\boldsymbol{\mu}\boldsymbol{\Gamma}_i)^T + \text{vec} \mathbf{t}_i^T)^T, \sum_{i=1}^n \mathbf{E}_{ii}^n \otimes \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Gamma}_i^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_i, h \right).$$

2.4 Invariance and nuisance parameters

In general, when a model contains nuisance parameters, the first step is to remove them. As in the matrix multivariate Gaussian model considered by Lele (1993), under an matrix multivariate elliptical model this objective is achieved through a simple transformation.

From (6)

$$\mathbf{X}_i \sim \mathcal{E}_{K \times D}(\boldsymbol{\mu}\boldsymbol{\Gamma}_i + \mathbf{t}_i, \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Gamma}_i^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_i, h), \quad i = 1, 2, \dots, n.$$

Recalling that $\mathbf{H}_K \mathbf{1}_K = \mathbf{0}_K$ and $\mathbf{1}_K^T \mathbf{H}_K = \mathbf{0}_K^T$, then, defining $\mathbf{X}_i^c = \mathbf{H}_K \mathbf{X}_i$, we have

$$\mathbf{X}_i^c \sim \mathcal{E}_{K \times D}^{(K-1), D}(\boldsymbol{\mu}^* \boldsymbol{\Gamma}_i, \boldsymbol{\Sigma}_K^* \otimes \boldsymbol{\Gamma}_i^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_i, h), \quad i = 1, 2, \dots, n. \quad (7)$$

where $\boldsymbol{\mu}^* = \mathbf{H}_K \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_K^* = \mathbf{H}_K \boldsymbol{\Sigma}_K \mathbf{H}_K$, $\mathbf{H}_K \mathbf{t}_i = \mathbf{H}_K \mathbf{1}_K \mathbf{a}_i^T = \mathbf{0}$ for all $i = 1, 2, \dots, n$, and $\boldsymbol{\mu}^*$ is such that its columns sum to zero, that is, it is a centered matrix.

Given that $K > D$ and that $\text{rank}(\boldsymbol{\Sigma}_K^*) = K - 1$, from Díaz-García and González-Farías (2005) and Díaz-García and Gutiérrez-Jáimez (2006) we have that

$$\mathbf{B}_i = \mathbf{X}_i^c (\boldsymbol{\Gamma}_i^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_i)^{-1} (\mathbf{X}_i^c)^T \sim \mathcal{GPW}_K^q(D, \boldsymbol{\Sigma}_K^*, \boldsymbol{\Sigma}_D, \boldsymbol{\Omega}, h), \quad i = 1, 2, \dots, n. \quad (8)$$

where

$$\boldsymbol{\Omega} = (\boldsymbol{\Sigma}_K^*)^{-} \boldsymbol{\mu}^* \boldsymbol{\Gamma}_i (\boldsymbol{\Gamma}_i^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_i)^{-1} \boldsymbol{\Gamma}_i^T (\boldsymbol{\mu}^*)^T = (\boldsymbol{\Sigma}_K^*)^{-} \boldsymbol{\mu}^* \boldsymbol{\Sigma}_D^{-1} (\boldsymbol{\mu}^*)^T.$$

$q = \min((K-1), D)$ and \mathbf{A}^{-} is any symmetric generalised inverse of \mathbf{A} such that $\mathbf{A} \mathbf{A}^{-} \mathbf{A} = \mathbf{A} = \mathbf{A}^T$. This is, \mathbf{B}_i has a *generalised singular pseudo-Wishart distribution*, which is independent of noise parameters.

Remark 2.3. Observe that \mathbf{B}_i can be write as

$$\mathbf{B}_i = \mathbf{X}_i^c (\boldsymbol{\Gamma}_i^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_i)^{-1} (\mathbf{X}_i^c)^T = \mathbf{X}_i^c \boldsymbol{\Gamma}_i^T \boldsymbol{\Sigma}_D^{-1} \boldsymbol{\Gamma}_i (\mathbf{X}_i^c)^T = \mathbf{Y}_i \boldsymbol{\Sigma}_D^{-1} \mathbf{Y}_i^T$$

where $\mathbf{Y}_i = \mathbf{X}_i^c \boldsymbol{\Gamma}_i^T$ and is such that

$$\mathbf{Y}_i \sim \mathcal{E}_{K \times D}^{(K-1), D}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}_K^* \otimes \boldsymbol{\Sigma}_D, h), \quad i = 1, 2, \dots, n.$$

In particular if $\boldsymbol{\Sigma}_D = \mathbf{I}_D$ and

$$\mathbf{X}_i^c = (\mathbf{X}_{1,i}^c | \mathbf{X}_{2,i}^c | \dots | \mathbf{X}_{D,i}^c)$$

with

$$\mathbf{X}_{d,i}^c \sim \mathcal{E}_K^{(K-1)}(\boldsymbol{\mu}^* \boldsymbol{\Gamma}_i \mathbf{e}_d^K, \boldsymbol{\Sigma}_K^*, h), \quad d = 1, 2, \dots, D; \quad i = 1, 2, \dots, n,$$

we have that,

$$\mathbf{B}_i = \mathbf{X}_i^c (\mathbf{X}_i^c)^T = \sum_{d=1}^D \mathbf{X}_{d,i}^c (\mathbf{X}_{d,i}^c)^T,$$

furthermore,

$$\mathbf{B}_i \sim \mathcal{GPW}_K^q(D, \boldsymbol{\Sigma}_K^*, \mathbf{I}_D, \boldsymbol{\Omega}, h), \quad i = 1, 2, \dots, n, \quad (9)$$

where $\boldsymbol{\Omega} = (\boldsymbol{\Sigma}_K^*)^{-} \boldsymbol{\mu}^* (\boldsymbol{\mu}^*)^T$.

Remark 2.4. The result in Lele (1993) is obtained as particular case of (9), with the difference that the matrix of noncentrality parameter in Lele (1993) is defined as $\boldsymbol{\mu}^*(\boldsymbol{\mu}^*)^T$ and we use $\boldsymbol{\Omega} = (\boldsymbol{\Sigma}_K^*)^{-1} \boldsymbol{\mu}^*(\boldsymbol{\mu}^*)^T$, notation used in (Muirhead, 1982, Definition 10.3.1, pp. 441-442).

In addition, defining \mathbb{X}^c as \mathbb{X} we have

$$\text{vec}(\mathbb{X}^c)^T = [\mathbf{I}_n \otimes (\mathbf{H}_k \otimes \mathbf{I}_D)] \text{vec} \mathbb{X}^T$$

hence, $\mathbb{X}^c = \mathbb{X}(\mathbf{H}_k \otimes \mathbf{I}_D)$. Now, observing that $(\mathbf{H}_k \otimes \mathbf{I}_D) \text{vec} \mathbf{t}_i^T = \mathbf{0}$, for all $i = 1, 2, \dots, n$. Then

$$\mathbb{X}^c \sim \mathcal{E}_{n \times KD}^{n, (K-1)D} \left(\sum_{i=1}^n \mathbf{e}_i^n \text{vec}^T(\boldsymbol{\mu}^* \boldsymbol{\Gamma}_i)^T, \sum_{i=1}^n \mathbf{E}_{ii}^n \otimes \boldsymbol{\Sigma}_K^* \otimes \boldsymbol{\Gamma}_i^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_i, h \right), \quad (10)$$

where $\boldsymbol{\Sigma}_K^* = \mathbf{H}_k \boldsymbol{\Sigma}_K \mathbf{H}_k$.

As in Lele (1993), assuming that $\boldsymbol{\Sigma}_D = \mathbf{I}_D$, and recalling that

$$\mathbf{I}_n = \sum_{i=1}^n \mathbf{E}_{ii}^n$$

we have

$$\mathbb{X}^c \sim \mathcal{E}_{n \times KD}^{n, (K-1)D} \left(\sum_{i=1}^n \mathbf{e}_i^n \text{vec}^T(\boldsymbol{\mu}^* \boldsymbol{\Gamma}_i)^T, \mathbf{I}_n \otimes \boldsymbol{\Sigma}_K^* \otimes \mathbf{I}_D, h \right).$$

3 Consistent estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_K$

Alternatively to the use of the Euclidean distance matrix showed in Lele (1993) with the aim to propose consistent estimations, we use directly the first two moments of the matrix \mathbf{B} with the same object.

When is considered a model where the perturbation of landmarks along the D axes are independent and identical to each other, formally we are assume that $\boldsymbol{\Sigma}_D = \mathbf{I}_D$ under a matrix multivariate Gaussian case. However, this same assumption is not to hold in matrix multivariate elliptical case. Under a matrix multivariate elliptical case is possible to consider two cases:

1. Independence and not correlation among landmarks and
2. Probabilistic dependence and not correlation among landmarks.

In both cases $\boldsymbol{\Sigma}_D = \mathbf{I}_D$ and the moments of matrix \mathbf{B} are different in each case.

Remark 3.1. Recall that under matrix multivariate elliptical distribution, only in the Gaussian case the not correlation and independence are equivalent. Then suppose that the vector $\mathbf{Z} = (z_1, z_2)^T$ has a bi-dimensional elliptical distribution and $\text{Cov}(\mathbf{Z}) = \mathbf{I}_2$ then z_1 and z_2 are independent if and only if \mathbf{Z} has a bi-dimensional Gaussian distribution. But if z_i , have a uni-dimensional elliptical distribution for $i = 1, 2$, and $\text{Var}(z_i) = 1$ and $\text{Cov}(z_1, z_2) = 0$, z_i , $i = 1, 2$, are not correlated and can be considered independent, see (Gupta and Varga, 1993, Section 6.2, p. 1) and (Fang, Kotz and Ng, 1990, Section 4.3, p. 105).

Summarising, given

$$\mathbf{B} = \mathbf{Y} \mathbf{Y}^T = \sum_{d=1}^D \mathbf{y}_d \mathbf{y}_d^T, \quad (11)$$

next, we find the first two moments of \mathbf{B} assuming that $\boldsymbol{\Sigma}_D = \mathbf{I}_D$, i.e. when \mathbf{y}_d : a) are not correlated and independent; and b) are not correlated and dependent.

3.1 Moments of \mathbf{B} under dependence

By completeness initially we assume that $\Sigma_D \neq \mathbf{I}_D$ and for convenience denote $\Sigma_D = \Theta$, $\Sigma_K^* = \Sigma$ and $\mu^* = \mu$.

With this goal in main, suppose that $\mathbf{Y} \sim \mathcal{E}_{K \times D}^{(K-1), D}(\mu, \Sigma \otimes \Theta, h)$, with

$$\mathbf{Y} = (\mathbf{y}_1 | \mathbf{y}_2 | \cdots | \mathbf{y}_D) \text{ and } \mu = (\mu_1 | \mu_2 | \cdots | \mu_D).$$

Observing that for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\text{vec } \mathbf{x}\mathbf{y}^T = \mathbf{y} \otimes \mathbf{x}$, $\mathbf{x}\mathbf{y}^T = \mathbf{x} \otimes \mathbf{y}^T = \mathbf{y}^T \otimes \mathbf{x}$ and thus, $\text{vec } \mathbf{y}\mathbf{y}^T \text{vec}^T \mathbf{y}\mathbf{y}^T = \mathbf{y} \otimes \mathbf{y}^T \otimes \mathbf{y} \otimes \mathbf{y}^T$, see Magnus and Neudecker (1979).

Theorem 3.1. *Let $\mathbf{Y} \sim \mathcal{E}_{K \times D}^{(K-1), D}(\mu, \Sigma \otimes \Theta, h)$. Then*

$$1. E(\text{vec } \mathbf{Y} \text{vec}^T \mathbf{Y}) = c_0(\Theta \otimes \Sigma) + \text{vec } \mu \text{vec}^T \mu,$$

$$2. \text{ and } E(\text{vec } \mathbf{Y} \text{vec } \mathbf{Y}^T \otimes \text{vec } \mathbf{Y} \text{vec } \mathbf{Y}^T) \text{ is}$$

$$\begin{aligned} &= \kappa_0[(\mathbf{I}_{(KD)^2} + \mathbf{K}_{KD})(\Theta \otimes \Sigma \otimes \Theta \otimes \Sigma) + \text{vec}(\Theta \otimes \Sigma) \text{vec}^T(\Theta \otimes \Sigma)] \\ &\quad + c_0(\mathbf{I}_{K^2} + \mathbf{K}_K)[\text{vec } \mu \text{vec}^T \mu \otimes (\Theta \otimes \Sigma) + (\Theta \otimes \Sigma) \otimes \text{vec } \mu \text{vec}^T \mu] \\ &\quad + c_0[\text{vec}(\Theta \otimes \Sigma)(\text{vec}^T \mu \mu^T) + (\text{vec}^T \mu \mu^T) \text{vec}(\Theta \otimes \Sigma)] \\ &\quad + \text{vec } \mu \text{vec}^T \mu \otimes \text{vec } \mu \text{vec}^T \mu, \end{aligned}$$

where \mathbf{K}_{KD} is the commutation matrix, see Magnus and Neudecker (1979), and $c_0 = E(u^2)$ and $3\kappa_0 = E(u^4)$, see (Gupta and Varga, 1993, p. 127),

$$E(u^2) = \frac{1}{i^2} \frac{d^2 \psi_U(t)}{dt^2} \Big|_{t=0} \text{ and } E(u^4) = \frac{1}{i^4} \frac{d^4 \psi_U(t)}{dt^4} \Big|_{t=0}.$$

Where $\psi_U(t) = \phi(t^2)$ is the characteristic function of univariate elliptical distribution. Some particular values of c_0 and κ_0 , are summarised on Table 1.

Proof. This is obtained differentiating (2) and observing that, see Díaz-García and Gutirrez Jimenez (1996),

$$\begin{aligned} E(\text{vec } \mathbf{Y} \otimes \text{vec}^T \mathbf{Y}) &= E(\text{vec } \mathbf{Y} \text{vec}^T \mathbf{Y}) \\ &= \frac{1}{i^2} \frac{\partial^2 \psi_{\text{vec } \mathbf{Y}}(\text{vec } \mathbf{T})}{\partial \text{vec } \mathbf{T} \partial \text{vec } \mathbf{T}^T} \Big|_{\text{vec } \mathbf{T}=0} \end{aligned}$$

and

$$\begin{aligned} E(\text{vec } \mathbf{Y} \otimes \text{vec } \mathbf{Y}^T \otimes \text{vec } \mathbf{Y} \otimes \text{vec } \mathbf{Y}^T) &= E(\text{vec } \mathbf{Y} \text{vec } \mathbf{Y}^T \otimes \text{vec } \mathbf{Y} \text{vec } \mathbf{Y}^T) \\ &= \frac{1}{i^4} \frac{\partial^4 \psi_{\text{vec } \mathbf{Y}}(\text{vec } \mathbf{T})}{\partial \text{vec } \mathbf{T} \partial \text{vec } \mathbf{T}^T \partial \text{vec } \mathbf{T} \partial \text{vec } \mathbf{T}^T} \Big|_{\text{vec } \mathbf{T}=0}. \end{aligned}$$

□

Now, given

$$\mathbf{B} = \mathbf{Y}\mathbf{Y}^T = \sum_{d=1}^D \mathbf{y}_d \mathbf{y}_d^T,$$

we have

$$E(\mathbf{B}) = E(\mathbf{Y}\mathbf{Y}^T) = E\left(\sum_{d=1}^D \mathbf{y}_d \mathbf{y}_d^T\right) = \sum_{d=1}^D E(\mathbf{y}_d \mathbf{y}_d^T).$$

Table 1: Particular values of c_0 and κ_0 .

Distribution	c_0	κ_0
Multiuniforme ^a	1	$\frac{1}{3}$
Gaussian ^b	1	1
Kotz ^c	$\frac{\Gamma\left[\frac{2N+1}{2s}\right]}{r^{1/s}\Gamma\left[\frac{2N-1}{2s}\right]}$	$\frac{\Gamma\left[\frac{2N+3}{2s}\right]}{3r^{2/s}\Gamma\left[\frac{2N-1}{2s}\right]}$
t^d	$\frac{m}{m-2}$	$\frac{m^2}{(m-2)(m-4)}$
Pearson Type II ^e	$\frac{1}{2m+3}$	$\frac{1}{(2m+3)(2m+5)}$
Pearson type VII ^f	$\frac{m}{2N-3}$	$\frac{m^2}{(2N-3)(2N-5)}$

^aFrom (Fang, Kotz and Ng, 1990, Theorem 3.3, p. 72).

^bFrom (Gupta and Varga, 1993, Remark 3.2.2, p. 125).

^cFrom (Nadarajah, 2003), where $r, s > 0$ and $2N + 1 > 2$.

^dFrom (Gupta and Varga, 1993, p. 128), or (Fang, Kotz and Ng, 1990, p. 88), where $m > 0$.

^eFrom (Fang, Kotz and Ng, 1990, Section 3.4.2, p. 89), where $m > -1$.

^fFrom (Fang, Kotz and Ng, 1990, Section 3.3.4, p. 84), where $N > 1/2$, $m > 0$.

And remembering that for $\mathbf{Y} \in \Re^{K \times D}$, in general

$$\text{Cov}(\text{vec } \mathbf{Y}) = E(\text{vec } \mathbf{Y} \text{vec}^T \mathbf{Y}) - E(\text{vec } \mathbf{Y})E(\text{vec}^T \mathbf{Y}).$$

Therefore

$$\begin{aligned}
\text{Cov}(\text{vec } \mathbf{B}) &= \text{Cov}(\text{vec}(\mathbf{Y}\mathbf{Y}^T)) = \text{Cov}\left(\sum_{d=1}^D \text{vec}(\mathbf{y}_d \mathbf{y}_d^T)\right) \\
&= E\left[\left(\sum_{d=1}^D \text{vec}(\mathbf{y}_d \mathbf{y}_d^T)\right)\left(\sum_{s=1}^D \text{vec}^T(\mathbf{y}_s \mathbf{y}_s^T)\right)\right] \\
&\quad - E\left(\sum_{d=1}^D \text{vec}(\mathbf{y}_d \mathbf{y}_d^T)\right)E\left(\sum_{d=1}^D \text{vec}^T(\mathbf{y}_d \mathbf{y}_d^T)\right) \\
&= \left[\sum_{d=1}^D \sum_{s=1}^D E(\mathbf{y}_d \mathbf{y}_s^T \otimes \mathbf{y}_d \mathbf{y}_s^T)\right] \\
&\quad - \text{vec}\left(\sum_{d=1}^D E(\mathbf{y}_d \mathbf{y}_d^T)\right)\text{vec}^T\left(\sum_{s=1}^D E(\mathbf{y}_s \mathbf{y}_s^T)\right) \tag{12}
\end{aligned}$$

Then, we need to find $E(\mathbf{y}_d \mathbf{y}_d^T)$ and $E(\mathbf{y}_d \mathbf{y}_s^T \otimes \mathbf{y}_d \mathbf{y}_s^T)$. These moments are obtained in the following result.

Theorem 3.2. Assume that $\mathbf{Y} \sim \mathcal{E}_{K \times D}^{(K-1), D}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}, h)$, with

$$\mathbf{Y} = (\mathbf{y}_1 | \mathbf{y}_2 | \cdots | \mathbf{y}_D) \text{ and } \boldsymbol{\mu} = (\boldsymbol{\mu}_1 | \boldsymbol{\mu}_2 | \cdots | \boldsymbol{\mu}_D),$$

and $\boldsymbol{\Theta} = (\theta_{ds})$. Then

$$1. E(\mathbf{y}_d \mathbf{y}_d^T) = c_0 \theta_{dd} \boldsymbol{\Sigma} + \boldsymbol{\mu}_d \boldsymbol{\mu}_d^T.$$

2. And

$$\begin{aligned} E(\mathbf{y}_d \mathbf{y}_s^T \otimes \mathbf{y}_d \mathbf{y}_s^T) &= \kappa_0 \theta_{ds}^2 [(\mathbf{I}_{K^2} + \mathbf{K}_K)(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) + \text{vec } \boldsymbol{\Sigma} \text{vec}^T \boldsymbol{\Sigma}] \\ &\quad + c_0 \theta_{ds} [(\mathbf{I}_{K^2} + \mathbf{K}_K)(\boldsymbol{\mu}_d \boldsymbol{\mu}_s^T \otimes \boldsymbol{\Sigma} + \boldsymbol{\Sigma} \otimes \boldsymbol{\mu}_d \boldsymbol{\mu}_s^T)] \\ &\quad + c_0 \theta_{ds} [\text{vec } \boldsymbol{\Sigma} \text{vec}^T \boldsymbol{\mu}_d \boldsymbol{\mu}_s^T + \text{vec } \boldsymbol{\mu}_d \boldsymbol{\mu}_s^T \text{vec}^T \boldsymbol{\Sigma}] \\ &\quad + \boldsymbol{\mu}_d \boldsymbol{\mu}_s^T \otimes \boldsymbol{\mu}_d \boldsymbol{\mu}_s^T. \end{aligned}$$

Proof. The results is obtained as consequence of Theorem 3.1 observing that $\mathbf{y}_d = \mathbf{Y} \mathbf{e}_d^D$, then

$$\begin{aligned} E(\mathbf{y}_d \mathbf{y}_d^T) &= E(\text{vec } \mathbf{y}_d \text{vec}^T \mathbf{y}_d) = E(\text{vec } \mathbf{Y} \mathbf{e}_d^D \text{vec}^T \mathbf{Y} \mathbf{e}_d^D) \\ &= (\mathbf{e}_d^{D^T} \otimes \mathbf{I}_K) E(\text{vec } \mathbf{Y} \text{vec}^T \mathbf{Y}) (\mathbf{e}_d^D \otimes \mathbf{I}_K) \\ &= (\mathbf{e}_d^{D^T} \otimes \mathbf{I}_K) (c_0 (\boldsymbol{\Theta} \otimes \boldsymbol{\Sigma}) + \text{vec } \boldsymbol{\mu} \text{vec}^T \boldsymbol{\mu}) (\mathbf{e}_d^D \otimes \mathbf{I}_K) \\ &= c_0 \theta_{dd} \boldsymbol{\Sigma} + \boldsymbol{\mu}_d \boldsymbol{\mu}_d^T. \end{aligned}$$

This least result is obtained noting that, $\text{vec } \mathbf{ABC} = (\mathbf{C}^T \otimes \mathbf{B}) \text{vec } \mathbf{B}$, $a \otimes \mathbf{A} = a\mathbf{A}$ and $(\mathbf{A} \otimes \mathbf{D})(\mathbf{B} \otimes \mathbf{E})(\mathbf{C} \otimes \mathbf{F}) = (\mathbf{ABC} \otimes \mathbf{DEF})$. Similarly,

$$E(\mathbf{y}_d \mathbf{y}_s^T \otimes \mathbf{y}_d \mathbf{y}_s^T) = \mathbf{R}^T E(\text{vec } \mathbf{Y} \text{vec}^T \mathbf{Y} \otimes \text{vec } \mathbf{Y} \text{vec}^T \mathbf{Y}) \mathbf{R}_1$$

with $\mathbf{R}^T = (\mathbf{e}_d^{D^T} \otimes \mathbf{I}_K) \otimes (\mathbf{e}_d^{D^T} \otimes \mathbf{I}_K)$ and $\mathbf{R}_1 = (\mathbf{e}_s^D \otimes \mathbf{I}_K) \otimes (\mathbf{e}_s^D \otimes \mathbf{I}_K)$. The desired result is obtained observing that: for $\mathbf{A} \in \mathbb{R}^{n \times s}$ and $\mathbf{B} \in \mathbb{R}^{m \times t}$, $\mathbf{K}_{mn}(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{B} \otimes \mathbf{A}) \mathbf{K}_{ts}$ and that $\mathbf{K}_{mm} \equiv \mathbf{K}_m$ see Magnus and Neudecker (1979). \square

Consider the following definition.

Definition 3.1. Let $\mathbf{A} \in \mathbb{R}^{p \times q}$ such that

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1n} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{m1} & \mathbf{A}_{m2} & \cdots & \mathbf{A}_{mn} \end{pmatrix}, \quad \mathbf{A}_{ij} \in \mathbb{R}^{r \times s}$$

with, $mr = p$ and $ns = q$, then

$$\bigoplus_{i,j}^{m,n} \mathbf{A} = \sum_{i=1}^m \sum_{j=1}^n \mathbf{A}_{ij} \in \mathbb{R}^{r \times s}.$$

If $m = n$ then, $\bigoplus_{i,j}^{m,m} \equiv \bigoplus_{i,j}^m$.

In addition, let $\mathbf{A} = (\mathbf{A}_{ij})$ and $\mathbf{B} = (\mathbf{B}_{ij})$ partitioned matrices. Then if \odot denotes the Khatri-Rao product, see (Rao, 1973, p.30),

$$\mathbf{A} \odot \mathbf{B} = (\mathbf{A}_{ij} \otimes \mathbf{B}_{ij})_{ij}.$$

In particular, note that if $\mathbf{C} = (c_{ij})$, then

$$\mathbf{C} \odot \mathbf{A} = (c_{ij} \mathbf{A}_{ij})_{ij}.$$

Moreover,

$$\bigoplus_{i,j} (\mathbf{C} \odot \mathbf{A}) = \sum_i \sum_j (c_{ij} \mathbf{A}_{ij})_{ij}.$$

Theorem 3.3. Suppose that $\mathbf{Y} \sim \mathcal{E}_{K \times D}^{(K-1), D}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}, h)$, with

$$\mathbf{Y} = (\mathbf{y}_1 | \mathbf{y}_2 | \cdots | \mathbf{y}_D) \text{ and } \boldsymbol{\mu} = (\boldsymbol{\mu}_1 | \boldsymbol{\mu}_2 | \cdots | \boldsymbol{\mu}_D).$$

And define

$$\mathbf{B} = \mathbf{Y}\mathbf{Y}^T = \sum_{d=1}^D \mathbf{y}_d \mathbf{y}_d^T.$$

Then

$$E(\mathbf{B}) = c_0 \text{tr}(\boldsymbol{\Theta}) \boldsymbol{\Sigma} + \boldsymbol{\mu} \boldsymbol{\mu}^T.$$

And

$$\begin{aligned} \text{Cov}(\text{vec } \mathbf{B}) &= (\mathbf{I}_{K^2} + \mathbf{K}_K) \left\{ \kappa_0 \text{tr}(\boldsymbol{\Theta}^2) (\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) \right. \\ &\quad + c_0 \left[\bigoplus_{i,j}^D (\boldsymbol{\Theta} \odot \text{vec } \boldsymbol{\mu} \text{vec}^T \boldsymbol{\mu}) \otimes \boldsymbol{\Sigma} \right. \\ &\quad \left. \left. + \boldsymbol{\Sigma} \otimes \bigoplus_{i,j}^D (\boldsymbol{\Theta} \odot \text{vec } \boldsymbol{\mu} \text{vec}^T \boldsymbol{\mu}) \right] \right\} \\ &\quad + [\kappa_0 \text{tr}(\boldsymbol{\Theta}^2) - c_0^2 \text{tr}^2(\boldsymbol{\Theta})] \text{vec } \boldsymbol{\Sigma} \text{vec}^T \boldsymbol{\Sigma} \\ &\quad + c_0 \left\{ \text{vec } \boldsymbol{\Sigma} \text{vec}^T \bigoplus_{i,j}^D (\boldsymbol{\Theta} \odot \text{vec } \boldsymbol{\mu} \text{vec}^T \boldsymbol{\mu}) \right. \\ &\quad \left. + \text{vec } \bigoplus_{i,j}^D (\boldsymbol{\Theta} \odot \text{vec } \boldsymbol{\mu} \text{vec}^T \boldsymbol{\mu}) \text{vec}^T \boldsymbol{\Sigma} \right. \\ &\quad \left. + \text{tr}(\boldsymbol{\Theta}) [\text{vec } \boldsymbol{\Sigma} \text{vec}^T \boldsymbol{\mu} \boldsymbol{\mu}^T + \text{vec } \boldsymbol{\mu} \boldsymbol{\mu}^T \text{vec}^T \boldsymbol{\Sigma}] \right\}. \end{aligned}$$

Proof. This is a consequence of (11), (12), Definition 3.1 and Theorem 2. \square

Corollary 3.1. In Theorem 3.3 assume that $\boldsymbol{\Theta} = \mathbf{I}_D$. Then

$$E(\mathbf{B}) = D c_0 \boldsymbol{\Sigma} + \boldsymbol{\mu} \boldsymbol{\mu}^T.$$

And

$$\begin{aligned} \text{Cov}(\text{vec } \mathbf{B}) &= (\mathbf{I}_{K^2} + \mathbf{K}_K) \left\{ D \kappa_0 (\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) + c_0 [\boldsymbol{\mu} \boldsymbol{\mu}^T \otimes \boldsymbol{\Sigma} + \boldsymbol{\Sigma} \otimes \boldsymbol{\mu} \boldsymbol{\mu}^T] \right\} \\ &\quad + D [\kappa_0 - D c_0^2] \text{vec } \boldsymbol{\Sigma} \text{vec}^T \boldsymbol{\Sigma} \\ &\quad + (1 - D) c_0 [\text{vec } \boldsymbol{\Sigma} \text{vec}^T \boldsymbol{\mu} \boldsymbol{\mu}^T + \text{vec } \boldsymbol{\mu} \boldsymbol{\mu}^T \text{vec}^T \boldsymbol{\Sigma}]. \end{aligned}$$

In univariate case, when $\boldsymbol{\mu} = \mathbf{0}$, these results were obtained in general and a particular cases in (Gupta and Varga, 1993, Theorem 3.2.13 and Example 3.2.1), with a several minor errors. In particular, for general case they write $D^2(\kappa_0 - c_0^2)$ and for matrix multivariate T distribution they write $D(\kappa_0 - c_0^2)$, with $D = n - 1$, instead of $D(\kappa_0 - D c_0^2)$.

3.2 Moments of \mathbf{B} under independence

Let \mathbf{Y} and $\boldsymbol{\mu}$ such that

$$\mathbf{Y} = (\mathbf{y}_1 | \mathbf{y}_2 | \cdots | \mathbf{y}_D) \text{ and } \boldsymbol{\mu} = (\boldsymbol{\mu}_1 | \boldsymbol{\mu}_2 | \cdots | \boldsymbol{\mu}_D),$$

where $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_D$ are independent and

$$\mathbf{y}_d \sim \mathcal{E}_K^{(K-1)}(\boldsymbol{\mu}_d, \theta_{dd} \boldsymbol{\Sigma}; h),$$

and by independence, $\text{Cov}(\mathbf{y}_d, \mathbf{y}_s) = \mathbf{0}$, for $d \neq s = 1, 2, \dots, D$.

Given

$$\mathbf{B} = \mathbf{Y}\mathbf{Y}^T = \sum_{d=1}^D \mathbf{y}_d \mathbf{y}_d^T,$$

we have

$$E(\mathbf{B}) = E(\mathbf{Y}\mathbf{Y}^T) = E\left(\sum_{d=1}^D \mathbf{y}_d \mathbf{y}_d^T\right) = \sum_{d=1}^D E(\mathbf{y}_d \mathbf{y}_d^T).$$

And under assumption that \mathbf{y}_d , $d = 1, 2, \dots, D$ are independent,

$$\begin{aligned} \text{Cov}(\text{vec } \mathbf{B}) &= \text{Cov}(\text{vec}(\mathbf{Y}\mathbf{Y}^T)) = \text{Cov}\left(\sum_{d=1}^D \text{vec}(\mathbf{y}_d \mathbf{y}_d^T)\right) \\ &= \sum_{d=1}^D \text{Cov}(\text{vec}(\mathbf{y}_d \mathbf{y}_d^T)) = \sum_{d=1}^D \text{Cov}(\mathbf{y}_d \otimes \mathbf{y}_d). \end{aligned}$$

Then, we need to find $E(\mathbf{y}_d \mathbf{y}_d^T)$ and

$$\begin{aligned} \text{Cov}(\mathbf{y}_d \otimes \mathbf{y}_d) &= E((\mathbf{y}_d \otimes \mathbf{y}_d)(\mathbf{y}_d \otimes \mathbf{y}_d)^T) - E(\mathbf{y}_d \otimes \mathbf{y}_d) E(\mathbf{y}_d \otimes \mathbf{y}_d)^T \\ &= E(\mathbf{y}_d \mathbf{y}_d^T \otimes \mathbf{y}_d \mathbf{y}_d^T) - E(\text{vec } \mathbf{y}_d \mathbf{y}_d^T) E(\text{vec}^T \mathbf{y}_d \mathbf{y}_d^T). \end{aligned}$$

These results are obtained in the following

Corollary 3.2. Let $\mathbf{y}_d \sim \mathcal{E}_K^{(K-1)}(\boldsymbol{\mu}_d, \theta_{dd}\boldsymbol{\Sigma}; h)$, $d = 1, 2, \dots, D$, where $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_D$ are independent. Then

1. $E(\mathbf{y}_d \mathbf{y}_d^T) = c_0 \theta_{dd} \boldsymbol{\Sigma} + \boldsymbol{\mu}_d \boldsymbol{\mu}_d^T$.
2. And $\text{Cov}(\mathbf{y}_d \otimes \mathbf{y}_d) = \text{Cov}(\text{vec } \mathbf{y}_d \mathbf{y}_d^T)$

$$\begin{aligned} &= (\mathbf{I}_{K^2} + \mathbf{K}_K) \{ \kappa_0 \theta_{dd}^2 (\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) + c_0 \theta_{dd} [\boldsymbol{\mu}_d \boldsymbol{\mu}_d^T \otimes \boldsymbol{\Sigma} + \boldsymbol{\Sigma} \otimes \boldsymbol{\mu}_d \boldsymbol{\mu}_d^T] \} \\ &\quad + \theta_{dd}^2 (\kappa_0 - c_0^2) \text{vec } \boldsymbol{\Sigma} \text{vec}^T \boldsymbol{\Sigma}. \end{aligned}$$

Proof. It is follows from Theorem 3.2, taking $d = s$. □

Theorem 3.4. Suppose that $\mathbf{y}_d \sim \mathcal{E}_K^{(K-1)}(\boldsymbol{\mu}_d, \theta_{dd}\boldsymbol{\Sigma}; h)$, $d = 1, 2, \dots, D$, with

$$\mathbf{Y} = (\mathbf{y}_1 | \mathbf{y}_2 | \dots | \mathbf{y}_D) \text{ and } \boldsymbol{\mu} = (\boldsymbol{\mu}_1 | \boldsymbol{\mu}_2 | \dots | \boldsymbol{\mu}_D),$$

and let $\boldsymbol{\Theta} = \text{diag}(\theta_{11}, \theta_{22}, \dots, \theta_{DD})$, and

$$\mathbf{B} = \mathbf{Y}\mathbf{Y}^T = \sum_{d=1}^D \mathbf{y}_d \mathbf{y}_d^T.$$

Then,

$$\begin{aligned} E(\mathbf{B}) &= c_0 \text{tr}(\boldsymbol{\Theta}) \boldsymbol{\Sigma} + \boldsymbol{\mu} \boldsymbol{\mu}^T \\ \text{Cov}(\text{vec } \mathbf{B}) &= (\mathbf{I}_{K^2} + \mathbf{K}_K) \{ \kappa_0 \text{tr}(\boldsymbol{\Theta}^2) (\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) \\ &\quad + c_0 \left[\left(\sum_{d=1}^D \theta_{dd} \boldsymbol{\mu}_d \boldsymbol{\mu}_d^T \right) \otimes \boldsymbol{\Sigma} + \boldsymbol{\Sigma} \otimes \left(\sum_{d=1}^D \theta_{dd} \boldsymbol{\mu}_d \boldsymbol{\mu}_d^T \right) \right] \} \\ &\quad + (\kappa_0 - c_0^2) \text{tr}(\boldsymbol{\Theta}^2) \text{vec } \boldsymbol{\Sigma} \text{vec}^T \boldsymbol{\Sigma} \end{aligned}$$

Proof. From Corollary 3.2,

$$\begin{aligned}
E(\mathbf{B}) &= E(\mathbf{Y}\mathbf{Y}^T) = E\left(\sum_{d=1}^D \mathbf{y}_d \mathbf{y}_d^T\right) = \sum_{d=1}^D E(\mathbf{y}_d \mathbf{y}_d^T) \\
&= \sum_{d=1}^D (c_0 \theta_{dd} \boldsymbol{\Sigma} + \boldsymbol{\mu}_d \boldsymbol{\mu}_d^T) \\
&= c_0 \text{tr}(\boldsymbol{\Theta}) \boldsymbol{\Sigma} + \sum_{d=1}^D \boldsymbol{\mu}_d \boldsymbol{\mu}_d^T = c_0 \text{tr}(\boldsymbol{\Theta}) \boldsymbol{\Sigma} + \boldsymbol{\mu} \boldsymbol{\mu}^T.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\text{Cov}(\text{vec } \mathbf{B}) &= \text{Cov}(\text{vec}(\mathbf{Y}\mathbf{Y}^T)) = \text{Cov}\left(\sum_{d=1}^D \text{vec}(\mathbf{y}_d \mathbf{y}_d^T)\right) \\
&= \sum_{d=1}^D \text{Cov}(\text{vec}(\mathbf{y}_d \mathbf{y}_d^T)).
\end{aligned}$$

from the desired result is obtained. \square

Now if $\boldsymbol{\Theta} = \mathbf{I}_D$, we have the following results.

Corollary 3.3. *Let $\mathbf{y}_d \sim \mathcal{E}_K^{(K-1)}(\boldsymbol{\mu}_d, \boldsymbol{\Sigma}; h)$, $d = 1, 2, \dots, D$, where $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_D$ are independent. Then*

1. $E(\mathbf{y}_d \mathbf{y}_d^T) = c_0 \boldsymbol{\Sigma} + \boldsymbol{\mu}_d \boldsymbol{\mu}_d^T$.
2. and $\text{Cov}(\mathbf{y}_d \otimes \mathbf{y}_d) = \text{Cov}(\text{vec } \mathbf{y}_d \mathbf{y}_d^T)$

$$\begin{aligned}
&= (\mathbf{I}_{K^2} + \mathbf{K}_K)[\kappa_0(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) + c_0(\boldsymbol{\mu}_d \boldsymbol{\mu}_d^T \otimes \boldsymbol{\Sigma} + \boldsymbol{\Sigma} \otimes \boldsymbol{\mu}_d \boldsymbol{\mu}_d^T)] \\
&\quad + (\kappa_0 - c_0^2) \text{vec } \boldsymbol{\Sigma} \text{vec}^T \boldsymbol{\Sigma},
\end{aligned}$$

Proof. It is immediately. \square

Theorem 3.5. *Suppose that $\mathbf{y}_d \sim \mathcal{E}_K^{(K-1)}(\boldsymbol{\mu}_d, \boldsymbol{\Sigma}; h)$, $d = 1, 2, \dots, D$, with*

$$\mathbf{Y} = (\mathbf{y}_1 | \mathbf{y}_2 | \dots | \mathbf{y}_D) \text{ and } \boldsymbol{\mu} = (\boldsymbol{\mu}_1 | \boldsymbol{\mu}_2 | \dots | \boldsymbol{\mu}_D),$$

and let

$$\mathbf{B} = \mathbf{Y}\mathbf{Y}^T = \sum_{d=1}^D \mathbf{y}_d \mathbf{y}_d^T.$$

Then,

$$\begin{aligned}
E(\mathbf{B}) &= D c_0 \boldsymbol{\Sigma} + \boldsymbol{\mu} \boldsymbol{\mu}^T \\
\text{Cov}(\text{vec } \mathbf{B}) &= (\mathbf{I}_{K^2} + \mathbf{K}_K)[D \kappa_0(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) + c_0(\boldsymbol{\mu} \boldsymbol{\mu}^T \otimes \boldsymbol{\Sigma} + \boldsymbol{\Sigma} \otimes \boldsymbol{\mu} \boldsymbol{\mu}^T)] \\
&\quad + D(\kappa_0 - c_0^2) \text{vec } \boldsymbol{\Sigma} \text{vec}^T \boldsymbol{\Sigma}
\end{aligned}$$

Proof. This is obtained from Theorem 3.4. \square

Corollary 3.4. *In particular if $\mathbf{Y} \sim \mathcal{N}_{K \times D}^{(K-1), D}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \mathbf{I}_D)$. Then, $c_0 = \kappa_0 = 1$, and thus*

$$\begin{aligned}
E(\mathbf{B}) &= D \boldsymbol{\Sigma} + \boldsymbol{\mu} \boldsymbol{\mu}^T \\
\text{Cov}(\text{vec } \mathbf{B}) &= (\mathbf{I}_{K^2} + \mathbf{K}_K)[D(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) + \boldsymbol{\mu} \boldsymbol{\mu}^T \otimes \boldsymbol{\Sigma} + \boldsymbol{\Sigma} \otimes \boldsymbol{\mu} \boldsymbol{\mu}^T].
\end{aligned}$$

3.3 Method-of-moments estimators

Returning to our notation, for which, rewrite, $\Theta = \Sigma_D$, $\Sigma = \Sigma_K^*$ and $\mu = \mu^*$.

Our target is to find the method-of-moments estimators of the parameter matrices Σ_K^* and μ^* . First, note that the first two sample moments estimators of \mathbf{B} are given by

$$\widetilde{E(\mathbf{B})} = \frac{1}{n} \sum_{i=1}^n \mathbf{B}_i = \bar{\mathbf{B}} = (\bar{b}_{ij}), \quad i, j = 1, \dots, K,$$

and

$$\text{Cov}(\widetilde{\text{vec } \mathbf{B}}) = \frac{1}{n} \sum_{i=1}^n (\text{vec } \mathbf{B}_i^T - \text{vec } \widetilde{E(\mathbf{B})}) (\text{vec } \mathbf{B}_i^T - \text{vec } \widetilde{E(\mathbf{B})})^T = \mathbf{S}.$$

where $\mathbf{S} = (s_{tr})$, $t, r = 1, 2, \dots, K^2$. In addition note that for $i \leq j$ and $\mathbf{M} = \mu^* \mu^{*T} = (m_{ij}) = \mathbf{M}^T$ and $\Sigma_K^* = (\sigma_{ij})$, we have

$$\begin{aligned} E(b_{ij}) &= E(\mathbf{e}_i^T \mathbf{B} \mathbf{e}_j) = \mathbf{e}_i^T E(\mathbf{B}) \mathbf{e}_j \\ &= \mathbf{e}_i^T (Dc_0 \Sigma_K^* + \mu^* \mu^{*T}) \mathbf{e}_j = Dc_0 \sigma_{ij} + m_{ij}, \end{aligned} \quad (13)$$

for independent and dependent cases.

3.3.1 Dependent case

Note that:

$$\begin{aligned} \text{Cov}(b_{ij}) &= \text{Cov}(\mathbf{e}_i^T \mathbf{B} \mathbf{e}_j) = \text{Cov}(\text{vec } \mathbf{e}_i^T \mathbf{B} \mathbf{e}_j) = \text{Cov}((\mathbf{e}_j \otimes \mathbf{e}_i)^T \text{vec } \mathbf{B}) \\ &= (\mathbf{e}_j \otimes \mathbf{e}_i)^T \text{Cov}(\text{vec } \mathbf{B}) (\mathbf{e}_j \otimes \mathbf{e}_i) \\ &= (\mathbf{e}_j \otimes \mathbf{e}_i)^T \{ (\mathbf{I}_{K^2} + \mathbf{K}_K) \{ D\kappa_0 (\Sigma_K^* \otimes \Sigma_K^*) \\ &\quad + c_0 [\mu^* \mu^{*T} \otimes \Sigma_K^* + \Sigma_K^* \otimes \mu^* \mu^{*T}] \} \\ &\quad + D [\kappa_0 - Dc_0^2] \text{vec } \Sigma_K^* \text{vec }^T \Sigma_K^* \\ &\quad + (1 - D)c_0 [\text{vec } \Sigma_K^* \text{vec }^T \mu^* \mu^{*T} \\ &\quad + \text{vec } \mu^* \mu^{*T} \text{vec }^T \Sigma_K^*] \} (\mathbf{e}_j \otimes \mathbf{e}_i). \end{aligned}$$

Observing that $(\mathbf{e}_j \otimes \mathbf{e}_i)^T \mathbf{K}_K = (\mathbf{e}_i \otimes \mathbf{e}_j)^T$ and that $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC} \otimes \mathbf{BD})$, for $i \leq j$, $i, j = 1, 2, \dots, K$

$$\begin{aligned} \text{Cov}(b_{ij}) &= D [\kappa_0 \sigma_{ii} \sigma_{jj} + (2\kappa_0 - c_0^2) \sigma_{ij}^2] \\ &\quad + c_0 [m_{jj} \sigma_{ii} + m_{ii} \sigma_{jj} + 2(2 - D)m_{ij} \sigma_{ij}]. \end{aligned} \quad (14)$$

From (13), by replacing $m_{ij} = \bar{b}_{ij} - Dc_0 \sigma_{ij}$ in (14) we have

$$\begin{aligned} \text{Cov}(b_{ij}) &= D(\kappa_0 - 2c_0^2) \sigma_{ii} \sigma_{jj} + D(2\kappa_0 - (1 + 2(2 - D))c_0^2) \sigma_{ij}^2 \\ &\quad + c_0 [\bar{b}_{jj} \sigma_{ii} + \bar{b}_{ii} \sigma_{jj} + 2(2 - D)\bar{b}_{ij} \sigma_{ij}]. \end{aligned} \quad (15)$$

Therefore equaling (15) to $s_{ij} = \widetilde{\text{Cov}(b_{ij})}$ we have:

Theorem 3.6. Assume that $\mathbf{B} \sim \mathcal{GPW}_K^q(D, \Sigma_K^*, \mathbf{I}_D, \Omega, h)$. Then, the method-of-moments estimators of Σ_K^* and $\mathbf{M} = \mu^* \mu^{*T}$ are given by the following exact expressions.

For $i = 1, 2, \dots, K$:

$$\tilde{\sigma}_{ii} = \frac{\sqrt{Q_{ii}^2 + 4Ps_{ii}} - Q_{ii}}{2P}, \quad (16)$$

where $Q_{ii}^2 + 4Ps_{ii} \geq 0$, $P = D(\kappa_0 - 2c_0^2) + D(2\kappa_0 - (1 + 2(2 - D))c_0^2)$, and $Q_{ii} = 2c_0(3 - D)\bar{b}_{ii}$.

$$\tilde{m}_{ii} = \bar{b}_{ii} - Dc_0\tilde{\sigma}_{ii}, \quad (17)$$

where $\tilde{\sigma}_{ii}$ has been previously found in (16).

If $P = 0$, then $\tilde{\sigma}_{ii} = s_{ii}/Q_{ii}$.

For $i < j$, $i = 1, \dots, (K - 1)$, $j = 2, \dots, K$:

$$\tilde{\sigma}_{ij} = \frac{\sqrt{(2 - D)^2 c_0^2 \bar{b}_{ij}^2 - R(T_{ij} - s_{ij}) - (2 - D)c_0 \bar{b}_{ij}}}{R}, \quad (18)$$

where $(2 - D)^2 c_0^2 \bar{b}_{ij}^2 - R(T_{ij} - s_{ij}) \geq 0$, $R = D(2\kappa_0 - (1 + 2(2 - D))c_0^2)$, and

$$T_{ij} = D(\kappa_0 - 2c_0^2)\tilde{\sigma}_{ii}\tilde{\sigma}_{jj} + c_0(\bar{b}_{jj}\tilde{\sigma}_{ii} + \bar{b}_{ii}\tilde{\sigma}_{jj}).$$

Here $\tilde{\sigma}_{ii}$ and $\tilde{\sigma}_{jj}$ were previously computed in (16).

$$\tilde{m}_{ij} = \bar{b}_{ij} - Dc_0\tilde{\sigma}_{ij}, \quad (19)$$

Denote the solution as

$$(\widetilde{\mathbf{M}}, \widetilde{\mathbf{\Sigma}}_K^*).$$

Note that $s_{ij} = \widehat{\text{Cov}(b_{ij})}$, this is s_{ij} are obtained from the diagonal of matrix $\mathbf{S} \in \Re^{K^2 \times K^2}$.

If $R = 0$, then $\tilde{\sigma}_{ij} = (s_{ij} - T_{ij}) / (2(2 - D)c_0\bar{b}_{ij})$.

Remark 3.2. Special attention must be payed on the constants P , Q_{ii} , R and T_{ij} , and the sign of the square root, according to the selected model and the sample statistics s_{ij} and \bar{b}_{ij} .

3.3.2 Independent case

For this case,

$$\begin{aligned} \text{Cov}(b_{ij}) &= \text{Cov}(\mathbf{e}_i^T \mathbf{B} \mathbf{e}_j) = \text{Cov}(\text{vec } \mathbf{e}_i^T \mathbf{B} \mathbf{e}_j) = \text{Cov}((\mathbf{e}_j \otimes \mathbf{e}_i)^T \text{vec } \mathbf{B}) \\ &= (\mathbf{e}_j \otimes \mathbf{e}_i)^T \text{Cov}(\text{vec } \mathbf{B})(\mathbf{e}_j \otimes \mathbf{e}_i) \\ &= (\mathbf{e}_j \otimes \mathbf{e}_i)^T \{(\mathbf{I}_{K^2} + \mathbf{K}_K)[D\kappa_0(\mathbf{\Sigma}_K^* \otimes \mathbf{\Sigma}_K^*) \\ &\quad + c_0(\boldsymbol{\mu}^* \boldsymbol{\mu}^{*T} \otimes \mathbf{\Sigma}_K^* + \mathbf{\Sigma}_K^* \otimes \boldsymbol{\mu}^* \boldsymbol{\mu}^{*T})] \\ &\quad + D(\kappa_0 - c_0^2) \text{vec } \mathbf{\Sigma}_K^* \text{vec}^T \mathbf{\Sigma}_K^*\} (\mathbf{e}_j \otimes \mathbf{e}_i). \end{aligned}$$

Hence

$$\text{Cov}(b_{ij}) = D[\kappa_0\sigma_{ii}\sigma_{jj} + (2\kappa_0 - c_0^2)\sigma_{ij}^2] + c_0(m_{jj}\sigma_{ii} + m_{ii}\sigma_{jj} + 2m_{ij}\sigma_{ij}). \quad (20)$$

From (13), by substituting $m_{ij} = s_{ij} - Dc_0\sigma_{ij}$ in (20) we have

$$\text{Cov}(b_{ij}) = D(\kappa_0 - 2c_0^2)\sigma_{ii}\sigma_{jj} + D(2\kappa_0 - 3c_0^2)\sigma_{ij}^2 + c_0[\bar{b}_{jj}\sigma_{ii} + \bar{b}_{ii}\sigma_{jj} + 2\bar{b}_{ij}\sigma_{ij}].$$

Summarising

Theorem 3.7. Assume that $\mathbf{y}_d \sim \mathcal{E}_K^{(K-1)}(\boldsymbol{\mu}_d, \mathbf{\Sigma}; h)$, independently, for $d = 1, 2, \dots, D$, such that

$$\mathbf{Y} = (\mathbf{y}_1 | \mathbf{y}_2 | \dots | \mathbf{y}_D) \text{ and } \boldsymbol{\mu} = (\boldsymbol{\mu}_1 | \boldsymbol{\mu}_2 | \dots | \boldsymbol{\mu}_D),$$

and let

$$\mathbf{B} = \mathbf{Y}\mathbf{Y}^T = \sum_{d=1}^D \mathbf{y}_d \mathbf{y}_d^T.$$

Then, the method-of-moments estimators of Σ_K^* and $\mathbf{M} = \boldsymbol{\mu}^* \boldsymbol{\mu}^{*T}$ are given by the following exact expressions.

For $i = 1, 2, \dots, K$:

$$\tilde{\sigma}_{ii} = \frac{\sqrt{Q_{ii}^2 + 4Ps_{ii}} - Q_{ii}}{2P}, \quad (21)$$

where $Q_{ii}^2 + 4Ps_{ii} \geq 0$, $P = D(3\kappa_0 - 5c_0^2)$, and $Q_{ii} = 4c_0\bar{b}_{ii}$.

$$\tilde{m}_{ii} = \bar{b}_{ii} - Dc_0\tilde{\sigma}_{ii}, \quad (22)$$

where $\tilde{\sigma}_{ii}$ has been previously found in (21).

If $P = 0$, then $\tilde{\sigma}_{ii} = s_{ii}/Q_{ii}$.

For $i < j$, $i = 1, \dots, (K-1)$, $j = 2, \dots, K$:

$$\tilde{\sigma}_{ij} = \frac{\sqrt{c_0^2\bar{b}_{ij}^2 - R(T_{ij} - s_{ij}) - c_0\bar{b}_{ij}}}{R}, \quad (23)$$

where $\bar{b}_{ij}^2 - R(T_{ij} - s_{ij}) \geq 0$, $R = D(2\kappa_0 - 3c_0^2)$ and

$$T_{ij} = D(\kappa_0 - 2c_0^2)\tilde{\sigma}_{ii}\tilde{\sigma}_{jj} + c_0\bar{b}_{jj}\tilde{\sigma}_{ii} + c_0\bar{b}_{ii}\tilde{\sigma}_{jj}.$$

Here $\tilde{\sigma}_{ii}$ and $\tilde{\sigma}_{jj}$ were previously computed in (21).

$$\tilde{m}_{ij} = \bar{b}_{ij} - Dc_0\tilde{\sigma}_{ij}, \quad (24)$$

Denote the solution as

$$(\widetilde{\mathbf{M}}, \widetilde{\Sigma}_K^*).$$

Note that $s_{ij} = \widetilde{\text{Cov}(\bar{b}_{ij})}$, this is s_{ij} are obtained from the diagonal of matrix $\mathbf{S} \in \Re^{K^2 \times K^2}$.

If $R = 0$, then $\tilde{\sigma}_{ij} = (s_{ij} - T_{ij}) / (2c_0\bar{b}_{ij})$.

Remark 3.3. Recall that the method-of-moments estimators are not uniquely defined. In addition, if instead of estimating the parameter θ , method-of-moments estimator of, say, $g(\theta)$ is desired, it can be obtained in several ways. One way would be to first find method-of-moments estimator, say $\tilde{\theta}$ of θ and then use $g(\tilde{\theta})$ as an estimator of $g(\theta)$. Alternatively, we can find the moments of function $g(\theta)$ and then apply the method of moments to find the method-of-moments estimator $\widetilde{g(\theta)}$ of $g(\theta)$. Estimators using either way are termed method-of-moments estimators and may be not be the same in both cases, see (Mood *et al*, 1974, Section 7.2.1, p.276).

The following result formalise the algorithm (Principal Coordinate Analysis, collected at Lele (1993)) for obtain $\boldsymbol{\mu}^*$, the estimated coordinates of the mean form (up to translation, rotation, and reflection transformations) using the method-of-moments estimator $\widetilde{\mathbf{M}}$.

Theorem 3.8. Let $\widetilde{\mathbf{M}}$ the method-of-moments estimator of $\mathbf{M} = \boldsymbol{\mu}^* \boldsymbol{\mu}^{*T}$ (for dependent or independent cases). Let $\widetilde{\mathbf{M}} = \mathbf{V}_1 \mathbf{L} \mathbf{V}_1^T$ is nonsingular part of its spectral decomposition, where \mathbf{V}_1 is a semiorthogonal matrix, $\mathbf{V}_1 \in \Re^{K \times D}$ i.e. $\mathbf{V}_1^T \mathbf{V}_1 = \mathbf{I}_D$ and $\mathbf{L} = \text{diag}(\lambda_1, \dots, \lambda_D)$, with D the rank of matrix $\widetilde{\mathbf{M}}$. Then the method-of-moments estimator of $\boldsymbol{\mu}^*$ is

$$\widetilde{\boldsymbol{\mu}}^* = \mathbf{V}_1 \mathbf{W},$$

where $\mathbf{W} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_D})$.

Proof. It is follow from Remark 3.3. \square

Theorem 3.9. Let $(\tilde{\boldsymbol{\mu}}^*, \tilde{\boldsymbol{\Sigma}}_K^*)$ the method-of-moments estimators of

$$(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}_K^*).$$

Then as $n \rightarrow \infty$

$$(\tilde{\boldsymbol{\mu}}^*, \tilde{\boldsymbol{\Sigma}}_K^*) \rightarrow (\boldsymbol{\mu}^*, \boldsymbol{\Sigma}_K^*) \quad \text{in probability.}$$

Proof. This follows from the consistency of the sample moments and the continuity of the function $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}_K^*)$ in $(E(\mathbf{B}), \text{Cov}(\text{vec } \mathbf{B}))$, see (Rao, 1973, Section 5d.1, p. 351). \square

4 Consistent estimation when $\boldsymbol{\Sigma}_D$ is a general non-negative definite matrix

Results in this section are motivated in the result obtained by Dutilleul (1999) under a matrix multivariate Gaussian distribution via the maximum likelihood estimation. We make an heuristic evaluation of the useful of these results in our approach based in method-of-moments estimation.

Our algorithm is based in the following modified expressions:

$$\tilde{\boldsymbol{\Sigma}}_D = \frac{1}{nK} \sum_{i=1}^n (\mathbf{X}_i^c - \tilde{\boldsymbol{\mu}}^*)^T (\tilde{\boldsymbol{\Sigma}}_K^*)^{-1} (\mathbf{X}_i^c - \tilde{\boldsymbol{\mu}}^*), \quad (25)$$

$$\tilde{\boldsymbol{\Sigma}}_K^* = \frac{1}{nD} \sum_{i=1}^n (\mathbf{X}_i^c - \tilde{\boldsymbol{\mu}}^*) \tilde{\boldsymbol{\Sigma}}_D^{-1} (\mathbf{X}_i^c - \tilde{\boldsymbol{\mu}}^*)^T. \quad (26)$$

ALGORITHM

INITIALISATION:

$$r = 0; \boldsymbol{\Sigma}_K^{*r} = \tilde{\boldsymbol{\Sigma}}_K^*; \boldsymbol{\Sigma}_D^r = \frac{1}{nK} \sum_{i=1}^n (\mathbf{X}_i^c - \tilde{\boldsymbol{\mu}}^*)^T (\boldsymbol{\Sigma}_K^{*r})^{-1} (\mathbf{X}_i^c - \tilde{\boldsymbol{\mu}}^*);$$

$r = r + 1$

$$\boldsymbol{\Sigma}_K^{*r+1} = \frac{1}{nD} \sum_{i=1}^n (\mathbf{X}_i^c - \tilde{\boldsymbol{\mu}}^*) (\boldsymbol{\Sigma}_D^r)^{-1} (\mathbf{X}_i^c - \tilde{\boldsymbol{\mu}}^*)^T;$$

$$\boldsymbol{\Sigma}_D^{r+1} = \frac{1}{nK} \sum_{i=1}^n (\mathbf{X}_i^c - \tilde{\boldsymbol{\mu}}^*)^T (\boldsymbol{\Sigma}_K^{*r})^{-1} (\mathbf{X}_i^c - \tilde{\boldsymbol{\mu}}^*);$$

WHILE

$$\|\boldsymbol{\Sigma}_D^{r+1} - \boldsymbol{\Sigma}_D^r\|_2 > \varepsilon_1 \text{ or } \|\boldsymbol{\Sigma}_K^{*r+1} - \boldsymbol{\Sigma}_K^{*r}\|_2 > \varepsilon_2,$$

REPEAT:

$$r = r + 1;$$

$$\boldsymbol{\Sigma}_K^{*r} = \boldsymbol{\Sigma}_K^{*r+1};$$

$$\boldsymbol{\Sigma}_D^r = \boldsymbol{\Sigma}_D^{r+1};$$

RECOMPUTE $\boldsymbol{\Sigma}_K^{*r+1}$ and $\boldsymbol{\Sigma}_D^{r+1}$.

SOLUTIONS ARE:

$$\tilde{\boldsymbol{\Sigma}}_K^* = \boldsymbol{\Sigma}_K^{*r}; \tilde{\boldsymbol{\Sigma}}_D = \boldsymbol{\Sigma}_D^r.$$

Where ε_1 and ε_2 define two infinitesimal positive quantities and $\|\cdot\|_2$ is the Euclidean norm, $\left(\|\mathbf{A}\|_2 = \sqrt{\text{tr}(\mathbf{A}\mathbf{A}^T)}\right)$.

Theorem 4.1. Let $(\tilde{\boldsymbol{\mu}}^*, \tilde{\boldsymbol{\Sigma}}_K^* \otimes \tilde{\boldsymbol{\Sigma}}_D)$ the method-of-moments estimators of $(\boldsymbol{\Sigma}_K^*, \boldsymbol{\mu}^* \otimes \boldsymbol{\Sigma}_D)$. Then as $n \rightarrow \infty$

$$(\tilde{\boldsymbol{\mu}}^*, \tilde{\boldsymbol{\Sigma}}_K^* \otimes \tilde{\boldsymbol{\Sigma}}_D) \rightarrow (\boldsymbol{\mu}^*, \boldsymbol{\Sigma}_K^* \otimes \boldsymbol{\Sigma}_D) \quad \text{in probability.}$$

Proof. This follows from Remark 3.3. \square

5 Estimation of the form difference

A detailed discussion of Euclidean Distance Matrix, matrix form, form difference and their probabilistic, geometrical, etc. properties may be found in Lele (1991, 1993). For your convenience, next we shall introduce some notation, although in general we adhere to standard notation forms.

Consider the following square symmetric matrix, know as Euclidean Distance Matrix:

$$\mathbf{F}(\mathbf{X}) = \begin{pmatrix} 0 & d(1, 2) & \dots & d(1, K-1) & d(1, K) \\ d(2, 1) & 0 & \dots & d(2, K-1) & d(2, K) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ d(K, 1) & d(K, 2) & \dots & d(K, K-1) & 0 \end{pmatrix},$$

where $d(i, j)$ denotes the Euclidean distance between landmarks i and j , in shape theory such matrix is termed *form matrix*. Among others interesting properties of form matrix, Lele (1991) proves that $\mathbf{F}(\mathbf{X})$ is a maximal invariant under the group of transformations consisting of translation, rotation, and reflection. Therefor, $\mathbf{F}(\mathbf{X})$ retains all the relevant information about the form of an object.

Let $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ be n independent observation from population I and $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m$ be m independent observation from population II. Let the mean form of population I be $\boldsymbol{\mu}^{\mathbf{X}}$ with the corresponding form matrix $\mathbf{F}(\boldsymbol{\mu}^{\mathbf{X}})$ and corresponding parameters for population II be $\boldsymbol{\mu}^{\mathbf{Y}}$ and $\mathbf{F}(\boldsymbol{\mu}^{\mathbf{Y}})$. From Lele (1993) we have the following definition:

Definition 5.1. Form difference between population I and II is defined as

$$\mathbf{FDM}(\boldsymbol{\mu}^{\mathbf{X}}, \boldsymbol{\mu}^{\mathbf{Y}}) = \mathbf{F}(\boldsymbol{\mu}^{\mathbf{X}}) * \mathbf{F}(\boldsymbol{\mu}^{\mathbf{Y}})^{-H},$$

where $*$ denotes the Hadamard product, $0/0 = 0$ and \mathbf{A}^{-H} denotes the inverse of \mathbf{A} with respect to the Hadamard product, a formula for such inverse in terms of the usual product is given in Caro-Lopera et al. (2012).

From remark 3.3, the following theorem shows that the form difference between two populations can be estimated consistently when landmarks are perturbed dependently along each axis but independently or not correlated between the axes.

Theorem 5.1. Let $(\boldsymbol{\mu}^{\mathbf{X}}, \boldsymbol{\Sigma}_{K\mathbf{X}}^* \otimes \boldsymbol{\Sigma}_{D\mathbf{X}})$ and $(\boldsymbol{\mu}^{\mathbf{Y}}, \boldsymbol{\Sigma}_{K\mathbf{Y}}^* \otimes \boldsymbol{\Sigma}_{D\mathbf{Y}})$ be the parameters for the two populations. If $\boldsymbol{\Sigma}_{D\mathbf{X}} = \boldsymbol{\Sigma}_{D\mathbf{Y}} = \mathbf{I}_D$, then

$$\widetilde{\mathbf{FDM}}(\tilde{\boldsymbol{\mu}}^{\mathbf{X}}, \tilde{\boldsymbol{\mu}}^{\mathbf{Y}}) = \tilde{\mathbf{F}}(\tilde{\boldsymbol{\mu}}^{\mathbf{X}}) * \tilde{\mathbf{F}}(\tilde{\boldsymbol{\mu}}^{\mathbf{Y}})^{-H} \Rightarrow \mathbf{FDM}(\boldsymbol{\mu}^{\mathbf{X}}, \boldsymbol{\mu}^{\mathbf{Y}}) \quad \text{in probability.}$$

Theorem 5.2. Let $(\boldsymbol{\mu}^{\mathbf{X}}, \boldsymbol{\Sigma}_{K\mathbf{X}}^* \otimes \boldsymbol{\Sigma}_{D\mathbf{X}})$ and $(\boldsymbol{\mu}^{\mathbf{Y}}, \boldsymbol{\Sigma}_{K\mathbf{Y}}^* \otimes \boldsymbol{\Sigma}_{D\mathbf{Y}})$ be the parameters for the two populations. Then

$$\widetilde{\mathbf{FDM}}(\tilde{\boldsymbol{\mu}}^{\mathbf{X}}, \tilde{\boldsymbol{\mu}}^{\mathbf{Y}}) = \tilde{\mathbf{F}}(\tilde{\boldsymbol{\mu}}^{\mathbf{X}}) * \tilde{\mathbf{F}}(\tilde{\boldsymbol{\mu}}^{\mathbf{Y}})^{-H} \rightarrow \mathbf{FDM}(\boldsymbol{\mu}^{\mathbf{X}}, \boldsymbol{\mu}^{\mathbf{Y}}) \quad \text{in probability.}$$

6 Example

The mouse vertebra problem was originally studied in the Gaussian case by Dryden and Mardia (1998) (see also Mardia and Dryden (1989)). A further analysis under elliptical models was implemented by Díaz-García and Caro-Lopera (2012b). The experiment considers the second thoracic vertebra T2 of two groups of mice: large and small. The mice are selected and classified according to large or small body weight; in this case, the sample consists of 23, 23 and 30 large, small and control bones, respectively. The vertebrae are digitised and summarised in six mathematical landmarks which are placed at points of high curvature, see figure 1; they are symmetrically selected by measuring the extreme positive and negative curvature of the bone. See Dryden and Mardia (1998) for more details. The shape difference analysis among the three groups is quite solved by a different approaches. However the correlation structure among landmarks requires more analysis; strong assumptions about those relations are usually considered because the complex exact shape distribution and a non existence theory for estimation for such invariant functions.

More than an example, this landmark data is highly valuable for a correlation structure analysis because the symmetry of the vertebra, certainly suggest a priori a non isotropic model. The control group is also useful for comparisons and correctness.

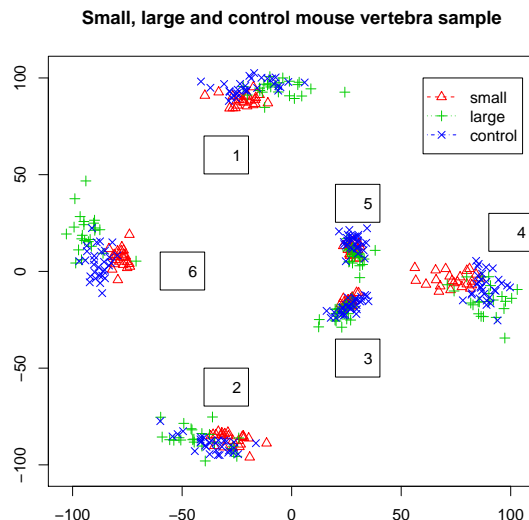


Figure 1: Mouse vertebra sample

Theorems 3.6 and 3.8 can be easily implemented for a number of models. We focus on the main novelty (Theorem 3.6) and Kotz type model (including Gaussian) which is very flexible and meaningful for various values of the parameters r, s and N , see appendix.

First of all we illustrate Theorem 3.7 under six different models with independent landmarks. Moment-method estimates of mean shape by using the common Gaussian model is shown in figure 2, in this case the estimate are complete unrealistic, as we expect, given the assumption of independence of landmarks. However, if we consider more complex models based on independence, the estimation tends to be more similar to the structure suggested by the sample. The addressed evolution from Kotz 1 to Kotz 5 is depicted in figures 3 to 7.

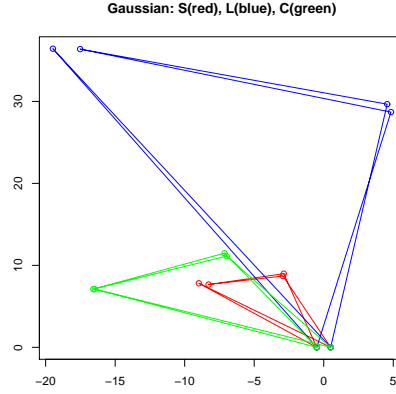


Figure 2: Moment method estimates under independence: Gaussian model

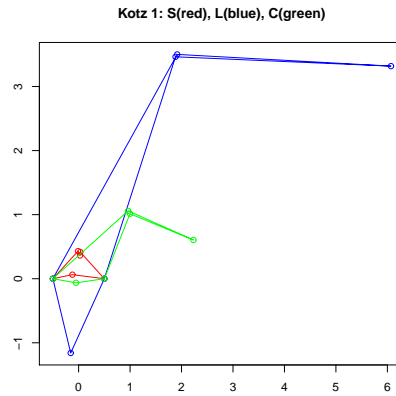


Figure 3: Moment method estimates under independence: Kotz 1 model

An heuristic behavior is noted, the lack of dependence in the Gaussian model, and its unrealistic moment method estimates, it seems to be improved by considering a more robust Kotz type model even with landmark dependence. The literature has studied this artificial data by the independent Gaussian case, so, given that no expert have set this assumption we can get further into more robust analysis and advance in some selection criteria, but if we have an experiment modeled by the independent Gaussian case according to the opinion of an expert in the field, we must follow that law and the further analysis, based on landmark dependence and elliptical families, that we provide next, cannot be implemented in such cases.

In the artificial mice data, we now can focus on the dependent case and the moment method estimators of Theorem 3.6, given that the Gaussian case is out of any consideration, then we have to study, for example, other Kotz models. In order to illustrate the important effect of landmark dependence we consider the simplest Kotz model after Gaussian, when $N = 2$, $r = 1/2$ and $s = 1$, which is referred as Kotz 1 model, and we compare the perfor-

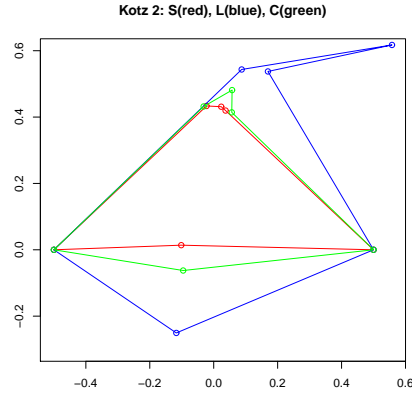


Figure 4: Moment method estimates under independence: Kotz 2 model

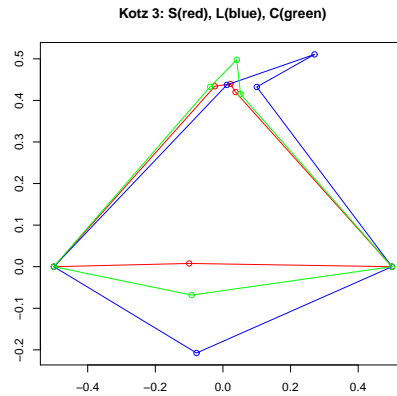


Figure 5: Moment method estimates under independence: Kotz 3 model

mance of Theorem 3.6 with another mean shape estimations. Table 2 provides comparisons among mean shape estimates of the small group, they include mean shape by moments of Theorem 3.6, the mean shape by Frechet method (see Kent (1992)), and Bookstein method (see Bookstein (1986)); certainly the estimations are truly similar. Note also that Kotz 1 law with independent landmarks provided a bad moment method estimator of mean shape, but the same model under the expected and realistic dependence reveals similarity with more complex mean shape estimators derived by standard shape theories, see figures 3 and 8, respectively.

The exact formula for the moments estimation (Theorem 3.6) also agrees with the previous conclusions in literature about strong difference in Gaussian mean shape between the small (S) and large (L) groups. Figure 8 also shows the mean shape estimation of the control (C) group. As we expect, the control group must tend to show strong symmetry among landmarks, by "averaging" in some sense the small and large estimates.

Different types of Kotz distribution have also modeled the sample, they correspond to

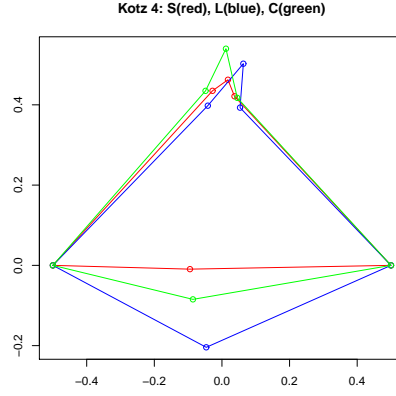


Figure 6: Moment method estimates under independence: Kotz 4 model

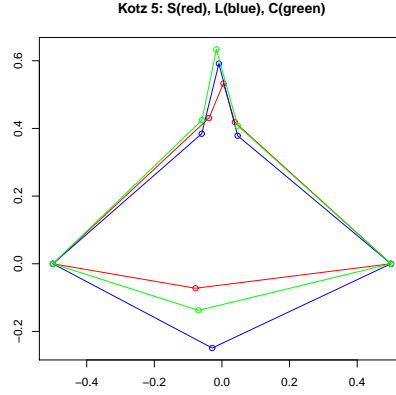


Figure 7: Moment method estimates under independence: Kotz 5 model

Table 2: Estimation for the mean shape for the small group by Theorem 3.6 (Kotz 1), Frechet (F), and Bookstein (B).

Th. 6, $\bar{\mu}_1$	Th. 6, $\bar{\mu}_2$	B. $\bar{\mu}_1$	B. $\bar{\mu}_2$	F. $\bar{\mu}_1$	F. $\bar{\mu}_2$
-0.5	0	-0.5	0	-0.5	0
0.5	0	0.5	0	0.5	0
0.084507028	0.3301634	0.08469746	0.2933430	0.08490820	0.2924684
0.014836162	0.6957339	0.01215768	0.5613175	0.01245608	0.5589496
-0.073397569	0.3394693	-0.06874750	0.2991278	-0.06869796	0.2982314
-0.005026754	-0.2184060	-0.02502185	-0.3041418	-0.02512807	-0.3044915

the denoted models Kotz 1, Kotz 2, Kotz 3, Kotz 4 and Kotz 5, with parameters $N = 2, s = 1, r = 1/2$; $N = 3, s = 1, r = 1/2$; $N = 2, s = 2, r = 1/2$; $N = 2, s = 3, r = 1/2$ and $N = 20, s = 20, r = 1/2$, respectively. Technical details about the generalised singular Pseudo-Wishart distributions and particular Kotz Pseudo-Wishart distributions referred in this example, can be seen in the appendix. The corresponding mean shapes estimates were computed, but for reasons of space, we only show the results of the Kotz 5 model (suggested

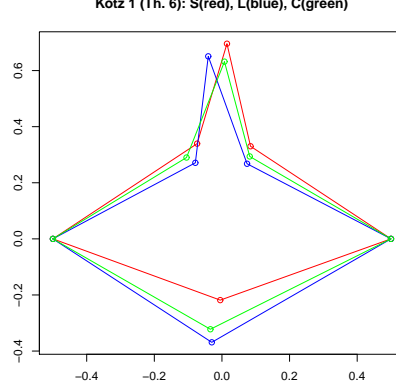


Figure 8: Moment method estimates under dependence: Kotz 1 model

by the preceding independent results and certain selection criteria that we will propose later) see figure 9.

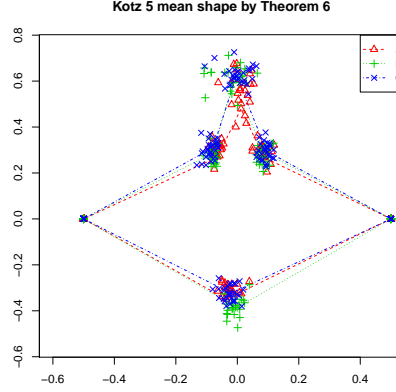


Figure 9: Mean shape estimates for large, small and control groups under the Kotz 5 model

Now we apply the algorithm for a consistent estimation when Σ_D is a general non-negative definite matrix, under the Kotz 5 law. For the routine propose in Section 4 we have fixed $\varepsilon_1 = \varepsilon_2 = 0.000005$ in the three groups small, large and control, then we found that the number of iteration to reach the addressed tolerance is 57, 53 and 61, respectively.

For the small group the estimated covariance matrices are given next (here the associated correlation matrix ρ of Σ is provided, for the sake of interpretation, recall that $\rho = (\text{diag}(\Sigma))^{-\frac{1}{2}} \Sigma (\text{diag}(\Sigma))^{-\frac{1}{2}}$):

$$\hat{\rho}_K^* = \begin{pmatrix} 1.0000000 & -0.88123087 & -0.45286210 & -0.0947470 & 0.3167573 & 0.1049686 \\ -0.8812309 & 1.00000000 & -0.01931031 & -0.3859582 & -0.7246992 & 0.3741399 \\ -0.4528621 & -0.01931031 & 1.00000000 & 0.9267571 & 0.6990041 & -0.9323209 \\ -0.0947470 & -0.38595825 & 0.92675709 & 1.0000000 & 0.9113738 & -0.9979133 \\ 0.3167573 & -0.72469917 & 0.69900414 & 0.9113738 & 1.0000000 & -0.9087033 \\ 0.1049686 & 0.37413987 & -0.93232089 & -0.9979133 & -0.9087033 & 1.0000000 \end{pmatrix};$$

and

$$\tilde{\rho}_D = \begin{pmatrix} 1.00000000 & -0.1305434 \\ -0.1305434 & 1.00000000 \end{pmatrix}.$$

For the large group the estimated correlation matrices are:

$$\tilde{\rho}_K^* = \begin{pmatrix} 1.00000000 & -0.65790585 & -0.49909037 & -0.05610956 & -0.07744806 & 0.42376772 \\ -0.65790585 & 1.00000000 & 0.06499537 & -0.44441269 & -0.29572585 & 0.03327717 \\ -0.49909037 & 0.06499537 & 1.00000000 & 0.42647926 & 0.65775361 & -0.76574318 \\ -0.05610956 & -0.44441269 & 0.42647926 & 1.00000000 & 0.65493138 & -0.73668655 \\ -0.07744806 & -0.29572585 & 0.65775361 & 0.65493138 & 1.00000000 & -0.79064351 \\ 0.42376772 & 0.03327717 & -0.76574318 & -0.73668655 & -0.79064351 & 1.00000000 \end{pmatrix};$$

and

$$\tilde{\rho}_D = \begin{pmatrix} 1.00000000 & -0.2080039 \\ -0.2080039 & 1.00000000 \end{pmatrix}.$$

Meanwhile in the control group the estimated correlation matrices are:

$$\tilde{\rho}_K^* = \begin{pmatrix} 1.00000000 & -0.6179802 & -0.6137305 & -0.3968700 & 0.05198526 & 0.5036286 \\ -0.61798019 & 1.00000000 & -0.1036402 & -0.4037052 & -0.70811391 & 0.2660900 \\ -0.61373047 & -0.1036402 & 1.00000000 & 0.7604036 & 0.50488092 & -0.8386367 \\ -0.39687003 & -0.4037052 & 0.7604036 & 1.00000000 & 0.64810504 & -0.9587295 \\ 0.05198526 & -0.7081139 & 0.5048809 & 0.6481050 & 1.00000000 & -0.6427126 \\ 0.50362856 & 0.2660900 & -0.8386367 & -0.9587295 & -0.64271257 & 1.00000000 \end{pmatrix};$$

and

$$\tilde{\rho}_D = \begin{pmatrix} 1.00000000 & 0.1048453 \\ 0.1048453 & 1.00000000 \end{pmatrix}.$$

The three groups reveal almost null correlation among axes, but some important correlation among landmarks, as we expect from the pseudo-symmetry of the bones. The estimates in the small and large groups detects the main landmarks responsible for the mean shape difference, meanwhile in the control case the estimates tends to follow the main contribution of large or small differentiating landmarks as we expect.

In a similar way we have run the routines with the same tolerance $\varepsilon_1 = \varepsilon_2 = 0.000005$ for the models Kotz 1, to Kotz 4; they reached the stability between 50 to 70 iterations in the three groups, and similar conclusions about the almost null correlation among axes and strong correlation among landmarks were found in the models. We will not show the estimates of each Kotz type, but have provided the results for the model Kotz 5 type, for reasons that we will explain later when the "best" model is selected under certain criteria; it was also suggested by the independent case analysis.

For a selection model criteria, the control group plays a fundamental role, in this case we just need to look for the law which obtains the minimum coefficient of variation when the small and large groups are compared with the control one; the analysis also must consider the distance between small and the large group relative to the mean with controls. We apply non-Euclidian distance between covariance matrix, a technique due to Dryden et al. (2009). The method is appropriate for meaningful correlation matrices, in this case it is performed only for $\tilde{\Sigma}_K^*$, because $\tilde{\Sigma}_D$ certainly ratifies in all the models that no correlation among axes is observed. In Tables 3 and 4, $K1, \dots, K5$, s , l , c , stand for Kotz 1, ..., Kotz 5, small, large and control, respectively.

Tables 3 and 4 shows all the pairwise covariance distances, in particular, the percentage variation coefficient is presented in parenthesis. We are searching for models which reflect

Table 3: Selection model criteria.

	K1l	K1c	K2s	K2l	K2c	K3s	K3l	K3c
K1s	12.9	8.7(37)	12.8	15.9	14.1	11.8	15.7	13.7
K1l		5.1(37)	10.6	6.1	8.4	10.8	5.1	8.5
K1c			9.6	9.1	8.9	9.2	8.6	8.7
K2s				11.0	11.0(14)	6.0	11.1	10.7
K2l					9.0(14)	12.0	1.5	9.1
K2c						11.6	8.8	0.9
K3s							12.1	11.2(15)
K3l								9.0(15)

Table 4: Selection model criteria.

	K4s	K4l	K4c	K5s	K5l	K5c
K1s	11.1	15.0	13.1	11.1	13.9	12.1
K1l	11.3	3.9	8.6	12.2	2.4	2.8
K1c	9.2	7.7	8.4	9.7	6.4	4.8
K2s	8.9	10.8	10.3	10.9	10.1	9.7
K2l	13.3	3.0	9.5	14.7	4.4	7.0
K2c	12.4	8.5	2.1	13.4	8.2	8.4
K3s	6.4	11.6	10.7	9.2	10.7	9.8
K3l	13.2	1.6	9.5	14.6	3.3	6.2
K3c	11.9	8.7	1.2	12.9	8.4	8.3
K4s		12.6	11.4(16)	6.8	11.6	10.6
K4l			9.1(16)	13.9	1.8	5.1
K4c				12.3	8.7	8.3
K5s					12.8	11.6(74)
K5l						3.6(74)

the role of the control group and separated the classes properly, the analysis must be complemented with mean shape estimates and a third criteria involving how the last ones are far from another accepted estimates, the Frechet mean shape for example. The addressed mean shape distance can be achieved by a number of approaches, see for example Kendall (1984).

Kotz 3 and 4 laws behave well with percentage variation coefficient, but the corresponding distance with the control group and the sample is too far to be realistic, specially with the small group. If we find the so called Riemannian distance among the moment method estimates and Frechet and Bookstein mean shape, we obtain the results of table 5:

Table 5: Model selection criteria.

	K2	K3	K4	K5	F.	B.
K1	0.274	0.236	0.211	0.180	0.113	0.112
K2		0.082	0.128	0.153	0.189	0.191
K3			0.048	0.078	0.131	0.133
K4				0.035	0.099	0.100
K5					0.067	0.068
F						0.002

The mean shape estimate based on a Kotz 5 model is very near to the estimates computed by Frechet and Bookstein (which are significantly similar), it also reflects good difference between the small and large; similar findings for the large and small group were computed, then collecting the results, we can propose Kotz type 5 model as a suitable law for modeling

this particular example. Note that the selection agrees with the conclusion proposed in the independent case.

It is important to note, that mathematical or statistical selection is just a suggestion for an experiment lacking of any prior assumption of the supporting distribution provided by an expert. In our case, literature shows no expert assumption about normality, in fact, this data full studied in Dryden and Mardia (1998) and the references therein, was traditionally set in the Gaussian theory in order to simplify computations and/or the use of the classes of exact distributions were not available at that time. However, if an experiment was sufficiently studied by an expert which the Gaussian model is truly normal, then the above selection of models, are out of significance; and given that moments-method estimates does not work under Gaussianity, as we have shown in this example, then the results presented here cannot be applied properly.

Now, at this stage, the conclusion about Kotz 5 model ratifies that non-Gaussian models explain better the three samples (an elliptical isotropic approach also verified this conclusion, see for example Díaz-García and Caro-Lopera (2012b)).

Once the model is selected, we are interested in application of Section 5, about estimation of mean form difference. In fact, we can go further by considering hypothesis testing for equality of the associated Euclidean Distance Matrices of two populations.

The methodology can be found in Lele and Richtsmeier (1991) and the references therein. We are interested in testing $H_0 : \mathbf{F}(\boldsymbol{\mu}^{\mathbf{X}}) = c\mathbf{F}(\boldsymbol{\mu}^{\mathbf{Y}})$, for some $c > 0$, where $\boldsymbol{\mu}^{\mathbf{X}}$ and $\boldsymbol{\mu}^{\mathbf{Y}}$ are the population mean shape. Based on a sample of objects \mathbf{X} 's and \mathbf{Y} 's, with corresponding estimated mean shapes $\tilde{\boldsymbol{\mu}}^{\mathbf{X}}$ and $\tilde{\boldsymbol{\mu}}^{\mathbf{Y}}$ obtained with the exact formula given in Theorem 3.6, we derive the form difference matrix $\mathbf{FDM}(\tilde{\boldsymbol{\mu}}^{\mathbf{X}}, \tilde{\boldsymbol{\mu}}^{\mathbf{Y}})$. This last matrix can be used for defining a number of suitable statistics for testing H_0 , however, Lele and Richtsmeier (1991) recommend the following:

$$T = \max_{i,j} FDM_{ij} / \min_{i,j} FDM_{ij},$$

where FDM_{ij} is the i, j -element of matrix \mathbf{FDM} . Note that if H_0 is true T is close to one. Moreover, T satisfies the desirable property of invariance under scaling, see Lele and Richtsmeier (1991) for more details.

The null distribution is difficult to obtain even in the simplest case of Gaussian, so we can obtain an empirical null distribution by using the well known bootstrap procedure, see Lele and Richtsmeier (1991) and the references therein. For similar samples of the current example, the above referred authors recommend a bootstrap of size 100.

Once the empirical distribution is obtain, a p-value, based on the upper tail of the observed statistics, rejects H_0 for small values near to 0.1.

Table 6 reports such tests for the Gaussian and Kotz 3 type models and the three pair comparisons of interest. We note that the usual Gaussian case, under the expected dependence condition of Theorem 3.6 malfunction and cannot detect the role of the control test, given a wrong conclusion. The selected model by covariance distances, separates as we expect the control group and gives a suitable p-value of certain difference, but it is not sufficient enough for concluding shape difference. This open an interesting discussion about the method based on coordinate free approach of Lele and Richtsmeier (1991), given that the quotient pairwise-element in definition of the matrix form difference is neglecting the whole matrix structure. Improving this aspect deserves a further work by defining a more robust matrix \mathbf{FDM} based on usual products than the very restrictive Hadamard product. Moreover, finding the corresponding exact distribution of $\mathbf{FDM}(\mathbf{X}, \mathbf{Y})$ can provide a promising null distribution which can model hypothesis testing efficiently.

Table 6: p-values of testing the equality of mean shape under different models and pairs of populations .

	Small-Large	Small-Control	Large-Control
Gaussian	0.00	0.00	0.00
Kotz 5	0.12	0.51	0.74

7 Conclusions

1. First, by replacing the Gaussian model for the elliptical model, an infinite range of possibilities in making an assumption of a model is opened, allowing to model a wide range of real situations, more or less heavy tails and more or less kurtosis than the Gaussian model.
2. Under this family of elliptical models is possible consistently estimate all parameters.
3. Notably, all these estimators are extremely easy to calculate.
4. Alternatively to the hypothesis assumed in subsections 2.3 and 2.4, an interesting alternative to investigate is: Assume that the joint distribution of $\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_n$ is

$$\mathbb{E} = (\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_n) \sim \mathcal{E}_{K \times nD}(\mathbf{0}, \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Sigma}_D \otimes \mathbf{I}_n, h),$$

where $\text{Cov}(\text{vec } \mathbb{E}^T) = \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Sigma}_D \otimes \mathbf{I}_n$. Then proceeding in a similar way and generalising to no central case the results in (Fang and Zhang, 1990, Eq. 3.4.14, p. 109) and (Gupta and Varga, 1993, Theorem 5.1.6, p. 170), we obtain that the joint distribution of $\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_n$ is

$$\mathbb{B} = (\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_n) \sim \mathcal{GPW}_{K,n} \left(\boldsymbol{\Sigma}_K^*, \frac{D}{2}, \frac{D}{2}, \dots, \frac{D}{2}, \boldsymbol{\Omega}, h \right),$$

where $\boldsymbol{\Omega} = (\boldsymbol{\Sigma}_K^*)^{-1} \boldsymbol{\mu}^* \boldsymbol{\Sigma}_D^{-1} \boldsymbol{\mu}^{*T}$. Noting that in this case it is implicitly assuming that the sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ is dependent.

5. Alternatively, and recalling that the method-of-moments estimators are not uniquely defined (see Remark 3.3) the method-of-moments estimator of $\boldsymbol{\Sigma}_D$ can be obtained from first two moments of \mathbf{B} too.

Acknowledgments

This article was written under the existing research agreement between the first author and the Universidad Autónoma Agraria Antonio Narro, Saltillo, México. F. Caro was supported by a research project of University of Medellín, Colombia.

References

- ARNOLD, S. F. (1981). Theory of linear models and multivariate analysis, John Wiley & Sons, Inc., New York.
- BOOKSTEIN, F. L. (1986). Size and shape spaces for landmark data in two dimensions (with discussion), *Stat. Sci.* **1** 181–242.

- CARO-LOPERA, F. C. AND DÍAZ-GARCÍA, J. A. (2012). Matrix Kummer-Pearson VII relation and polynomial pearson VII configuration density, *J. Iranian Stat. Soc.* **11**(2) 217–230.
- CARO-LOPERA, F. J., DÍAZ-GARCÍA, J. A., AND GONZÁLEZ-FARÍAS, G. (2009). Noncentral elliptical configuration density, *J. Multivariate Anal.* **101**(1) 32–43.
- CARO-LOPERA, F. J., DÍAZ-GARCÍA, J. A., AND GONZÁLEZ-FARÍAS, G. (2014). Inference in affine shape theory under elliptical models, *J. Korean Stat. Soc.* **43**(1) 67–77.
- CARO-LOPERA, F. J., LEIVA, V. AND BALAKRISHNAN, N. (2012). Connection between the Hadamard and matrix products with an application to matrix-variate Birnbaum-Saunders distributions, *J. Multivariate Anal.* **104** 126–139.
- DÍAZ-GARCÍA, J. A., AND CARO-LOPERA, F. J. (2012a). Generalised shape theory via SV decomposition I, *Metrika* **75**(4) 541–565.
- DÍAZ-GARCÍA, J. A., AND CARO-LOPERA, F. J. (2012b). Statistical theory of shape under elliptical models and singular value decompositions, *J. Multivariate Anal.* **103**(1) 77–92.
- DÍAZ-GARCÍA, J. A., AND CARO-LOPERA, F. J. (2013). Generalised shape theory via pseudo Wishart distribution, *Sankhyā A* **75**(2) 253–276.
- DÍAZ-GARCÍA, J. A., AND CARO-LOPERA, F. J. (2014). Statistical theory of shape under elliptical models via QR decomposition, *Statistics* **48**(2) 456–472.
- DÍAZ-GARCÍA, J. A. (1994). *Contributions to the theory of Wishart and multivariate elliptical distributions*, Ph.D. dissertation, Universidad de Granada, Spain, (in Spanish).
- DÍAZ-GARCÍA, J. A., AND R. GUTIÉRREZ JÍMEZ, R. (1996). *Matrix differential calculus and moments of a random matrix elliptical*, Serie Coleccin "Estadística Multivariable y Procesos Estocásticos". Universidad de Granada, España, (in Spanish).
- DÍAZ-GARCÍA, J. A., AND GONZÁLEZ-FARÍAS, G. (2005). Singular Random Matrix decompositions: Distributions, *J. Multivariate Anal.* **94**(1) 109–122.
- DÍAZ-GARCÍA, J. A., GUTIÉRREZ, R. J. AND RAMOS-QUIROGA, R. (2003). Size-and-shape cone, shape disk and configuration densities for the elliptical models, *Brazilian J. Prob. Stati.* **17** 135–146.
- DÍAZ-GARCÍA, J. A., AND GUTIÉRREZ-JÁIMEZ, R. (2006). Wishart and Pseudo-Wishart distributions under elliptical laws and related distributions in the shape theory context, *J. Statist. Plann. Inference* **136**(12) 4176–4193.
- DRYDEN, I. L. AND MARDIA, K. V. (1998). *Statistical shape analysis*, John Wiley and Sons, Chichester.
- DRYDEN, I. L., KOLOYDENKO, A., AND ZHOU D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging, *An. App. Statist.* **3** 1102–1123.
- DUTILLEUL, P. (1999). The mle algorithm for the matrix normal distribution, *J. Statist. Comput. Simul.* **64** 105–123.
- FANG, K. T. AND ZHANG, Y. T. (1990). *Generalized multivariate analysis*, Science Press, Beijing, Springer-Verlang.

- FANG, K. T., KOTZ, S., AND NG, K. W. (1990). *Symmetric multivariate and related distributions*, Chapman Hall, London.
- GOODALL, C. R. (1991). Procrustes methods in the statistical analysis of shape (with discussion), *J. Royal Stat. Soc. B* **53** 285–339.
- GOODALL, C. R., AND MARDIA, K. V. (1993). Multivariate aspects of shape theory, *Ann. Statist.* **21** 848–866.
- GUPTA, A. K., AND VARGA, T. (1993). *Elliptically contoured models in statistics*, Kluwer Academic Publishers, Dordrecht
- KENT D. G. (1984). Shape manifolds, Procrustean metrics and complex projective spaces, *Bulletin of the London Mathematical Society*, **16** 81–121.
- KENT J. T. (1992). *New directions in shape analysis*, In Mardia, K. V., editor, *The Art of Statistical Science*, 115–127. Wiley, Chichester.
- Khatri, C. G. (1968). Some results for the singular normal multivariate regression models, *Sankhyā A* **30** 267–280.
- KOEV, P., AND EDELMAN, A. (2006). The efficient evaluation of the hypergeometric function of a matrix argument, *Math. Comp.* **75** 833–846.
- LE, H. L., AND KENDALL, D. G. (1993). The Riemannian structure of Euclidean spaces: a novel environment for statistics, *Ann. Statist.* **21** 1225–1271.
- LELE, S. (1991). Some Comments on Coordinate Free and Scale Invariant Method in Morphometrics, *Am. J. Phys. Anthropol.* **85** 407–418.
- LELE, S. (1993). Euclidean distance matrix analysis (EDMA): Estimation of mean form and mean form difference, *Math. Geol.* **25**(5) 573–602.
- LELE, S., AND RICHTSMEIER, J. (1991). Euclidean distance matrix analysis: A coordinate free approach for comparing biological shapes using landmark data, *Amer. J. Phys. Anthropol.* **86** 415–427.
- LELE, S., AND RICHTSMEIER, J. (1990). Statistical models in morphometrics: Are they realistic? *Syst. Zool.* **39**(1) 60–69.
- MAGNUS, J. R., AND NEUDECKER, H. (1979). The commutation matrix: Some properties and applications, *Ann. Statist.* **7**(2) 381–394.
- MARDIA, K. V., AND DRYDEN, I. L. (1989) The Statistical Analysis of Shape Data, *Biometrika* **76** 271–281.
- MOOD, A. M., GRAYBILL, F. A., AND BOES, D. C. (1974). *Introduction to the theory statistics*, Third edition, McGraw-Hill Series in Probability and Statistics New York.
- MUIRHEAD, R. J. (1982). *Aspects of multivariate statistical theory*, Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.
- NADARAJAH, S. (2003). The Kotz-type distribution with applications, *Statistics* **37**(4) 341–358.
- RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*, John Wiley & Sons, New York, 1973.

RICHTSMEIER, J. T., DELEON, V. B., AND LELE, S. R. (2002). The Promise of Geometric Morphometrics, *Yearbook of Phis. Anthropol.* **45** (2002) 63–91.

WALKER, J. (2001). Ability of geometric morphometric methods to estimate a known covariance matrix, *Syst. Biol.* **49** 686–696.

A Particular generalised Pseudo-Wishart singular distributions

The following result is a particular case of the general result in Díaz-García and González-Farías (2005) or Díaz-García and Gutiérrez-Jáimez (2006), when Θ is non singular and the notation of this paper is assumed.

Theorem A.1 (Generalised singular Pseudo-Wishart distributions). *Assume that $\mathbf{Y} \sim \mathcal{E}_{K \times D}^{K-1,D}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \Theta, h)$, where h admits a power series expansion*

$$h(v + a) = \sum_{t=0}^{\infty} \frac{h^{(t)}(a)v^t}{t!}.$$

in \mathfrak{R} . Let, also, $q = \min(K - 1, D)$; then the density of $\mathbf{B} = \mathbf{Y}\Theta^{-1}\mathbf{Y}^T$ is given by

$$= \frac{\pi^{qD/2} |\mathbf{L}|^{(D-K-1)/2}}{\Gamma_q(\frac{1}{2}D) \left(\prod_{i=1}^{(K-1)} \lambda_i^{D/2} \right)} \sum_{t=0}^{\infty} \sum_{\kappa} \frac{h^{(2t)}(\text{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{B} + \boldsymbol{\Omega}))}{t!} \frac{C_{\kappa}(\boldsymbol{\Omega}\boldsymbol{\Sigma}^{-1}\mathbf{B})}{(\frac{1}{2}D)_{\kappa}} (d\mathbf{B}) \quad (27)$$

where $\mathbf{B} = \mathbf{W}_1 \mathbf{L} \mathbf{W}_1^T$, is the nonsingular spectral decomposition of \mathbf{B} with \mathbf{W}_1 a semiorthogonal matrix, i.e. $\mathbf{W}_1^T \mathbf{W}_1 = \mathbf{I}_q$, and $\mathbf{L} = \text{diag}(l_1, \dots, l_q)$; $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \boldsymbol{\mu}^T$, $(d\mathbf{B})$ is Hausdorff measure is defined in (Díaz-García and González-Farías, 2005, Section 5); λ_i , $i = 1, \dots, (K - 1)$, are nonnull eigenvalues of $\boldsymbol{\Sigma}$, and where $C_{\kappa}(\mathbf{A})$ are the zonal polynomials of \mathbf{A} corresponding to the partition $\kappa = (t_1, \dots, t_{\alpha})$ of t , with $\sum_1^{\alpha} t_i = t$; $(a)_{\kappa} = \prod_{j=1}^{\alpha} (a - (j - 1)/2)_{t_j}$, $(a)_t = a(a + 1) \cdots (a + t - 1)$, being the generalized hypergeometric coefficients and $\Gamma_s(a) = \pi^{s(s-1)/4} \prod_{j=1}^s \Gamma(a - (j - 1)/2)$ is the multivariate gamma function, see Muirhead (1982);

Corollary A.1 (Singular Pseudo-Wishart Gaussian distribution). *Let us suppose that $\mathbf{Y} \sim \mathcal{N}_{K \times D}^{K-1,D}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \Theta)$, and let $q = \min(K - 1, D)$; then the density of $\mathbf{B} = \mathbf{Y}\Theta^{-1}\mathbf{Y}^T$ is given by*

$$= C \text{etr} \left(-\frac{1}{2}(\boldsymbol{\Sigma}^{-1}\mathbf{B} - \boldsymbol{\Omega}) \right) {}_0F_1 \left(\frac{1}{2}D; \frac{1}{4}\boldsymbol{\Omega}\boldsymbol{\Sigma}^{-1}\mathbf{B} \right) (d\mathbf{B}), \quad (28)$$

with

$$C = \frac{\pi^{D(q-(K-1))/2} |\mathbf{L}|^{(D-K-1)/2}}{2^{D(K-1)/2} \Gamma_q[\frac{1}{2}D] \left(\prod_{i=1}^{K-1} \lambda_i^{D/2} \right)},$$

where ${}_0F_1(\cdot)$ is a hypergeometric function with a matrix argument, see (Muirhead, 1982, p. 258).

B Singular Pseudo-Wishart Kotz distribution.

Firs recall that the $K \times D$ random matrix \mathbf{X} is said to have a *singular matrix multivariate symmetric Kotz type distribution* with parameters $N, r, s \in \mathbb{R}$, $\boldsymbol{\mu} : K \times D$, $\boldsymbol{\Sigma} : K \times K$, of rank $K - 1$, $\boldsymbol{\Theta} : D \times D$ with $r > 0$, $s > 0$, $2N + (K - 1)D > 2$, $\boldsymbol{\Sigma} > \mathbf{0}$, and $\boldsymbol{\Theta} > \mathbf{0}$ if its density is

$$\frac{s r^{(2N+(K-1)D-2)/2s} \Gamma((K-1)D/2)}{\pi^{(K-1)D/2} \Gamma[(2N+(K-1)D-2)/2s] \left(\prod_{i=1}^{K-1} \lambda_i^{D/2} \right) |\boldsymbol{\Theta}|^{(K-1)/2}} \\ \times [\text{tr } \boldsymbol{\Theta}^{-1}(\mathbf{Y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu})]^{N-1} \exp \{ -r \text{tr}^s \boldsymbol{\Theta}^{-1}(\mathbf{Y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu}) \}.$$

When $T = s = 1$, and $R = 1/2$ we get the singular matrix variate gaussian distribution.

Note that particular singular Pseudo-Wishart distributions just depend on the general derivative $h^{(2t)}(\cdot)$ of the elliptical generator function; it seems a trivial fact, but the general formulae involves cumbersome expressions indexed by partitions, see Caro-Lopera et al. (2009). In the case of Kotz type distribution they derived the following expressions.

When $s = 1$, the Kotz type models and their general derivative simplify substantially. Thus, the following expressions applies for Gaussian, and the so called Kotz 1, Kotz 2, with parameters $N = 1, s = 1, r = 1/2$; $N = 2, s = 1, r = 1/2$; $N = 3, s = 1, r = 1/2$; respectively. The generator model is given by

$$h(y) = \frac{r^{N-1+(K-1)D/2} \Gamma[(K-1)D/2]}{\pi^{(K-1)D/2} \Gamma[N-1+(K-1)D/2]} y^{N-1} \exp\{-ry\},$$

And, the corresponding k -th derivative of h , follows from

$$\frac{d^k}{dy^k} y^{N-1} \exp[-ry],$$

which is given by

$$(-r)^k y^{N-1} \exp[-ry] \left\{ 1 + \sum_{v=1}^k \binom{k}{v} \left[\prod_{i=0}^{v-1} (N-1-i) \right] (-ry)^{-v} \right\},$$

where $k = 2t$.

For the remaining models of the example, the so termed Kotz 3, Kotz 4 and Kotz 5, with parameters $N = 2, s = 2, r = 1/2$; $N = 2, s = 3, r = 1/2$ and $N = 20, s = 20, r = 1/2$, respectively, the generator function is given by:

$$h(y) = \frac{s r^{(2N+(K-1)D-2)/2s} \Gamma[(K-1)D/2]}{\pi^{(K-1)D/2} \Gamma[(2N+(K-1)D-2)/2s]} y^{N-1} \exp(-ry^s)$$

meanwhile the required k -th derivative of h , follows from $\frac{d^k}{dy^k} \exp(-ry^s)$, which is given by

$$y^{T-1} e^{-Ry^s} \left\{ \sum_{\kappa \in P_k} \frac{k! (-R)^{\sum_{i=1}^k v_i} \prod_{j=0}^{k-1} (s-j)^{\sum_{i=j+1}^k v_i}}{\prod_{i=1}^k v_i! (i!)^{v_i}} y^{\sum_{i=1}^k (s-i)v_i} \right. \\ \left. + \sum_{m=1}^k \binom{k}{m} \left[\prod_{i=0}^{m-1} (T-1-i) \right] \right\}$$

$$\times \sum_{\kappa \in P_{k-m}} \frac{(k-m)!(-R)^{\sum_{i=1}^{k-m} v_i} \prod_{j=0}^{k-m-1} (s-j)^{\sum_{i=j+1}^{k-m} v_i}}{\prod_{i=1}^{k-m} v_i! (i!)^{v_i}} y^{\sum_{i=1}^{k-m} (s-i)v_i - m} \Bigg\},$$

where $\sum_{\kappa \in P_k}$ denotes the summation over all the partitions

$$\kappa = \left(k^{v_k}, (k-1)^{v_{k-1}}, \dots, 3^{v_3}, 2^{v_2}, 1^{v_1} \right)$$

of k , with $\sum_{i=1}^k i v_i = k$, i.e. κ is a partition of k consisting of v_1 ones, v_2 twos, v_3 threes, etc. It is important to quote that all the singular Pseudo-Wishart distributions associated with the above Kotz type kernels can be computed by some modifications of the algorithms provided by Koev and Edelman (2006) for the Gaussian case, see for example Díaz-García and Caro-Lopera (2013) and similar works of the authors on shape theory.